

Adaption multivariater Testmethoden für hochdimensionale Daten der funktionellen Bildgebung

Dissertation

zur Erlangung des akademischen Grades

**doctor rerum naturalium
(Dr. rer. nat)**

von Dipl.-Stat.(FH) Daniela Adolf

geb. am 04.11.1982 in Magdeburg

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: apl. Prof. Dr. Waltraud Kahle
apl. Prof. Dr. Siegfried Kropf

eingereicht am: 07.12.2011

Verteidigung am: 03.05.2012

Inhaltsverzeichnis

Symbolverzeichnis	III
Zusammenfassung	VI
Abstract	VII
1. Einleitung	1
2. Mathematische Grundlagen	5
2.1. Verteilungen	5
2.1.1. Multi- und matrixvariante Normalverteilung	5
2.1.2. Linkssphärische Matrixverteilung	7
2.2. Das klassische allgemeine lineare Modell	8
2.3. Testverfahren für das klassische allgemeine lineare Modell	10
2.3.1. Univariate Analysen	11
2.3.2. Multivariate Analysen	12
2.3.3. Permutationstests	19
2.3.4. Rotationstests	24
3. Modellbesonderheiten bei funktionellen MRT-Daten	28
3.1. Grundlagen der funktionellen Magnetresonanztomographie	28
3.2. Besonderheiten des allgemeinen linearen Modells für fMRT-Daten	32
3.2.1. Format der Beobachtungsmatrix	32
3.2.2. Aufbau der Designmatrix	33
3.2.3. Kovarianzstruktur der Beobachtungen	35
3.3. Notwendigkeit einer Anpassung an die zeitliche Korrelation	40
4. Etablierte Anpassungsstrategien für korrelierte Stichprobenelemente	44
4.1. Univariate Anpassungsstrategien für zeitlich korrelierte Daten	44
4.1.1. Satterthwaite-Approximation	44
4.1.2. Pre-Whitening	46
4.2. Analoge multivariate Anpassungsstrategien für zeitlich korrelierte Daten	47
4.2.1. Satterthwaite-Approximation	47
4.2.2. Pre-Whitening	47
4.3. Überlegungen zur Umsetzung der parametrischen Adaptionstrategien	48
4.3.1. Schätzung der Korrelationsstruktur	48
4.3.2. Andere parametrische Ansätze zur Schätzung der Kovarianzstruktur	54

5. Adaption von Permutationsverfahren für den Umgang mit korrelierten Daten	56
5.1. Blockpermutationen	56
5.2. Neue Blockpermutationen mit Random Shift	61
5.3. Auswahl geeigneter Blocklängen	63
6. Gegenüberstellung der Verfahren anhand simulierter Daten	71
6.1. Design der Simulationsstudie	71
6.2. Empirischer Fehler 1. Art	75
6.2.1. Ergebnisse der simulierten fMRT-Daten	75
6.2.2. Ergebnisse der simulierten Kontrollsznarien	83
6.3. Gütesimulationen	85
6.3.1. Güteresultate der simulierten fMRT-Daten	85
6.3.2. Güteresultate der simulierten Kontrollsznarien	97
7. Anwendung der Verfahren auf reale fMRT-Daten	100
7.1. Hintergrund der Studie zum Ultimatum Game	100
7.2. Beschreibung der Daten	102
7.3. Ergebnisse der multivariaten Auswertung	103
8. Resümee	111
Literaturverzeichnis	113
A. Totale Produktsummenmatrix bei Permutation	119
B. Erwartungswert des AR(1)-Korrelationskoeffizientenschätzers	122
C. Weitere Simulationsergebnisse für den Fehler 1. Art	123
D. Weitere Simulationsergebnisse für die Güte	128
Danksagung	130

Symbolverzeichnis

a	Skalar
\mathbf{a}_n	Vektor der Dimension n (Spaltenvektor, wenn nicht anders angegeben)
$\mathbf{A}_{n \times p}$	Matrix über n Zeilen und p Spalten
\mathbf{A}'	Transponierte von \mathbf{A}
\mathbf{A}^{-1}	Inverse von \mathbf{A}
\mathbf{A}^-	Moore-Penrose-Inverse von \mathbf{A}
$\hat{\mathbf{A}}$	Schätzer von \mathbf{A}
\mathbf{A}_π	permutierte Version von \mathbf{A} (zeilenweise Permutation)
\mathbf{A}_δ	rotierte Version von \mathbf{A}
$E(\mathbf{a})$	Erwartungswert von \mathbf{a}
$\text{Var}(\mathbf{a})$	Varianz von \mathbf{a}
$\text{Cov}(\mathbf{a}, \mathbf{b})$	Kovarianz von \mathbf{a} und \mathbf{b}
$\text{Rg}(\mathbf{A})$	Rang von \mathbf{A}
$\text{Spur}(\mathbf{A})$	Spur von \mathbf{A}
$ \mathbf{A} $	Determinante von \mathbf{A}
$\mathbf{A} \mathbf{B}$	\mathbf{A} bei gegebenem \mathbf{B}
$\min(a, b)$	Minimum von a und b
$\max(a, b)$	Maximum von a und b
$a!$	Fakultät von a
$\#(a < b)$	Anzahl der Fälle $a < b$
$\stackrel{d}{=}$	gleich in Verteilung
\otimes	Kronecker-Produkt
$*$	Faltungsoperator
$\mathbf{1}_n$	n -dimensionaler Spaltenvektor bestehend aus Einsen
\mathbf{c}	Kontrastgewichtsvektor der Dimension $s \times 1$
\mathbf{C}	Kontrastgewichtsmatrix der Dimension $s \times m$
\mathbf{d}	Gewichtsvektor der Dimension $p \times 1$
\mathbf{D}	Gewichtsmatrix der Dimension $p \times q$
F	F -Teststatistik
F_{FG_1, FG_2}	Quantil der F -Verteilung mit den Freiheitsgraden FG_1 und FG_2
g	Quadratsummenterm zur Berechnung der univariaten F -Teststatistik (residuale Abweichungen)
\mathbf{G}	Produktsummenmatrix der residualen Abweichungen

h	Quadratsummenterm zur Berechnung der univariaten F -Teststatistik (hypothesenbezogene Abweichungen)
\mathbf{H}	Produktsummenmatrix der hypothesenbezogenen Abweichungen
H_0	Nullhypothese
H_1	Alternativhypothese
\mathbf{I}_n	Identmatrix der Dimension n
k	Anzahl der Blöcke bei Blockpermutation
l	Blocklänge bei Blockpermutation
m	Dimension der Nullhypothese
n	Anzahl der Stichprobenvektoren / Anzahl der Zeilen der Beobachtungsmatrix
$N(\mu, \sigma^2)$	Normalverteilung mit Erwartungswert μ und Varianz σ^2
n_{sim}	Anzahl der Simulationen
n_δ	Anzahl zufälliger Rotationen
n_π	Anzahl zufälliger Permutationen
N_π	Anzahl aller möglichen Permutationen
p	Variablenanzahl / Anzahl der Spalten der Beobachtungsmatrix
P	P -Wert
P_{MC}	P -Wert bei Monte-Carlo-Simulationen
$\mathbf{Q}_G, \mathbf{Q}_H$	Projektionsmatrizen
q_{ij}	quadratischer euklidischer Abstand zwischen zwei Vektoren \mathbf{y}_i und \mathbf{y}_j
r_{ij}	Pearson'scher Korrelationskoeffizient zwischen zwei Vektoren \mathbf{y}_i und \mathbf{y}_j
\mathbf{R}	residuenbildende Matrix
\mathbb{R}^p	Menge aller reellen Zahlen im p -dimensionalen Raum
s	Anzahl der Regressoren
s_{ij}	Ähnlichkeitsmaß zwischen zwei Vektoren \mathbf{y}_i und \mathbf{y}_j
\mathbf{S}	Ähnlichkeitsmatrix
t	Teststatistik nach Student
t_{FG}	Quantil der t -Verteilung mit dem Freiheitsgrad FG
t_{ij}	Abstandsmaß zwischen zwei Vektoren \mathbf{y}_i und \mathbf{y}_j
\mathbf{T}	Abstandsmatrix
T_0	Teststatistik der Originalstichprobe
T_π	Teststatistik bei Permutation
T_δ	Teststatistik bei Rotation
U_{FG_1, FG_2, FG_3}	Quantil der Verteilung von Wilk's Λ mit den Freiheitsgraden FG_1 , FG_2 und FG_3
\mathbf{W}	totale Produktsummenmatrix der Dimension $p \times p$
\mathbf{X}	Designmatrix der Dimension $n \times s$
\mathbf{y}	Beobachtungsvektor der Dimension $n \times 1$
\mathbf{Y}	Beobachtungsmatrix der Dimension $n \times p$
\mathbf{Z}	Scorematrix der Dimension $n \times q$

α	Irrtumswahrscheinlichkeit
β	Parametervektor der Dimension $s \times 1$
\mathbf{B}	Parametermatrix der Dimension $s \times p$
γ	Kovarianzfunktion des AR(1)-Prozesses
Δ	Rotationsmatrix
ϵ	Residuenvektor der Dimension $n \times 1$
\mathbf{E}	Residuenmatrix der Dimension $n \times p$
λ	Eigenwert
Λ	Wilk's Λ
$\mathbf{\Lambda}$	Eigenwertmatrix
μ	Erwartungswert
ν	approximierter Freiheitsgrad der t -Verteilung
ρ	Korrelationskoeffizient
ρ_r	räumlicher Korrelationskoeffizient
\mathbf{P}	zeitliche Korrelationsmatrix der Dimension $n \times n$
σ^2	Varianz der Residuen
$\mathbf{\Sigma}$	räumliche Kovarianzmatrix der Dimension $p \times p$
τ	Lag / Abstand zweier Zeitpunkte im AR(1)-Modell
$\mathbf{\Psi}$	zeitliche Kovarianzmatrix der Dimension $n \times n$

Zusammenfassung

Ziel dieser Arbeit ist die Anwendung multivariater Verfahren auf Daten der funktionellen Magnetresonanztomographie (fMRT) zum Nachweis neuronaler Aktivität.

Bei fMRT-Aufnahmen eines Probanden handelt es sich um hochdimensionale Daten, die neben einer räumlichen Korrelation auch zeitliche Abhängigkeiten aufweisen. Die Anzahl der Variablen, der sogenannten Voxel, übersteigt dabei den Stichprobenumfang, welcher der Anzahl aufeinanderfolgender Messungen entspricht, zumeist erheblich.

Deshalb werden speziell sogenannte stabilisierte multivariate Testverfahren auf ihre Anwendbarkeit auf fMRT-Daten untersucht. Diese Verfahren basieren auf linkssphärischen Verteilungen und arbeiten für unabhängige Stichprobenvektoren auch im hochdimensionalen Fall exakt. Die Datenanalyse kann dabei auf Basis eines allgemeinen linearen Modells erfolgen. In fMRT-Untersuchungen sind die aufeinanderfolgenden Messungen jedoch miteinander korreliert, so dass die Tests auf diese Dateneigenschaft angepasst werden müssen. Standardmäßig wird für die Abhängigkeitsstruktur der Stichprobenvektoren ein autoregressiver Prozess erster Ordnung unterstellt.

Zur Adjustierung der multivariaten Testverfahren an diese zeitliche Korrelation werden einerseits zwei univariat etablierte metrische Methoden, eine Satterthwaite-Approximation und ein Pre-Whitening-Verfahren, für multivariate Daten adaptiert. Andererseits wird eine neue Adjustierungsmethode in Form einer speziellen Permutationsstrategie vorgeschlagen. Diese blockweise Permutation mit zufälliger Verschiebung berücksichtigt die Abhängigkeit der Stichprobenvektoren, ohne jedoch eine spezielle Annahme an die Korrelationsstruktur zu stellen.

Die Anwendung der unterschiedlichen multivariaten Teststatistiken in Kombination mit den Adjustierungsverfahren für zeitliche Korrelation wird mittels simulierter fMRT-Daten bezüglich der empirischen Fehler 1. und 2. Art untersucht. Anschließend werden die einzelnen Verfahren zusätzlich anhand realer Probandendaten einer fMRT-Studie miteinander verglichen.

In dieser Arbeit kann so gezeigt werden, dass die stabilisierten multivariaten Teststatistiken erfolgreich für den Nachweis von Hirnaktivität eingesetzt werden können und insbesondere bei schwachen Aktivierungen Vorteile gegenüber den univariaten, für multiples Testen adjustierten Analysen aufweisen. Zudem erweist sich das erarbeitete blockweise Permutationsverfahren als eine adäquate Alternative zur Adjustierung der Tests an die zeitliche Abhängigkeitsstruktur von fMRT-Daten.

Abstract

The intention of this thesis is to apply multivariate analyses to functional magnetic resonance imaging (fMRI) data in order to detect neuronal activation.

For an fMRI session with one subject, the measured signal response is a time series over a large amount of variables, so-called voxels, and the data are correlated in space and time. Usually the number of variables is much higher than the sample size.

Because of the high dimension, so-called stabilized multivariate test statistics are investigated for applicability to fMRI data. These tests are based on left-spherical distributions and deliver parametrically exact results even in a high-dimensional framework in case of independent sample vectors. Data analysis can be realized by using a general linear model. Since the consecutive measurements are correlated in fMRI data, the test methods have to be adapted to this dependence, for which a first-order autoregressive process is generally assumed.

To adjust the multivariate tests to the temporal correlation, two established metric methods are applied: a Satterthwaite approximation and a pre-whitening. Additionally, a new adjustment method in terms of a special permutation strategy is proposed. This block-wise permutation including a random shift accounts for the dependence of the sample vectors without the necessity of a specific assumption concerning the correlation structure.

The application of the different multivariate test statistics in combination with the three adjustment strategies for correlated samples is investigated for the empirical type I and type II errors by using simulated fMRI data. Subsequently, these methods are compared by means of real fMRI data.

This thesis reveals that stabilized multivariate test statistics can successfully be applied to detect neuronal activation. Especially in the case of low activation levels, these methods show advantages in comparison to univariate test statistics, which have to be adjusted for multiple testing. Furthermore, the proposed block-wise permutation strategy proves to be an adequate alternative for adjustment to the temporal correlation in fMRI data.

1. Einleitung

Die funktionelle Magnetresonanztomographie (fMRT) zählt in den Neurowissenschaften zu den etablierten Methoden zur Untersuchung des menschlichen Gehirns. Durch sie ist es möglich, die Verarbeitung von kognitiven Reizen den entsprechenden Hirnarealen zuzuordnen. Der Entwicklungsprozess dieses Verfahrens ist von Beginn an mit vielfältigen statistischen Überlegungen und Ansätzen verbunden, zu denen die vorliegende Arbeit einen Beitrag liefern möchte.

Die Magnetresonanztomographie ist ein nicht-invasives Bildgebungsverfahren, das die magnetischen Eigenschaften von Protonen zur Erzeugung eines Signals nutzt. Die Signalintensität ist abhängig von der Stärke des Magnetfelds sowie der Protonendichte. Diese wiederum ist ein Maß für die im Gewebe vorhandenen Wasserstoffkerne. Somit ist die Signalintensität letztlich gewebespezifisch (Schneider & Fink, 2007). Über eine dreidimensionale Zuordnung des Signals ist so eine detaillierte strukturelle Bildgebung möglich. Die hierbei referenzierten Volumenelemente, in denen die Signalstärken als Grauwerte dargestellt sind, werden als Voxel (volumetric pixel) bezeichnet.

Auch der Sauerstoffgehalt im Blut, konkret das Verhältnis von freiem und an Sauerstoff gebundenem Hämoglobin, hat Einfluss auf die Signalstärke. Da die Aktivierung von Neuronen im Gehirn mit einer lokalen Veränderung des Sauerstoffgehalts assoziiert ist, kann die neuronale Aktivität indirekt über die Änderung im sogenannten BOLD- (blood-oxygen-level-dependent) Signal gemessen werden (Ogawa *et al.*, 1990). Allerdings tritt diese Signaländerung zeitlich verzögert auf.

Um das BOLD-Signal zu analysieren und dadurch Aktivierungen in den einzelnen Voxeln nachzuweisen, sind Versuchsanordnungen und Analysemodelle erforderlich, die eine realistische Beschreibung der fMRT-Daten ermöglichen. Bei der Auswertung der Gehirnaktivität eines einzelnen Probanden, die im Fokus dieser Arbeit steht, handelt es sich in der fMRT um eine sogenannte first-level-Analyse. Hierbei hat sich die Anwendung des allgemeinen linearen Modells etabliert, in dem die wiederholten Messungen des BOLD-Signals als Stichprobenelemente und die einzelnen Voxel als Variablen betrachtet werden. Erst durch spezielle Paradigmen, in denen die Differenz des BOLD-Signals bei Aktivierung zu einem Signal ohne neuronale Aktivität bestimmt wird, kann die Veränderung im BOLD-Kontrast vom Rauschen im MR-Signal unterschieden werden. Denn die Aktivierungsdifferenz macht nur wenige Prozent des messbaren Gesamtsignals aus. Weiterhin wird eine bessere Unterscheidung dadurch ermöglicht, dass sich Aktivierungs- und Ruhebedingung wiederholt ablösen (Dössel, 2000). Die Reihenfolge der Reize sowie verschiedene weitere Einfluss- und Störfaktoren, wie eine Korrektur für die Bewegung des Probanden, fließen dabei in die Designmatrix des linearen Modells ein.

Die für diese Arbeit entscheidende Besonderheit bei den first-level-Analysen besteht darin, dass die Realisierung schnell aufeinanderfolgender Aufnahmen mit dem Problem der Abhängigkeit dieser gemessenen Werte einhergeht. Denn aufgrund der langsamen Veränderung des BOLD-Signals weisen zeitlich benachbarte fMRT-Aufnahmen eine Korrelation auf (Friston *et al.*, 2007). Da die klassische Statistik jedoch von unabhängigen Stichprobenelementen beziehungsweise -vektoren ausgeht, müssen die verwendeten Analyseverfahren an die Situation zeitlich korrelierter Daten angepasst werden.

Zudem sind Daten der funktionellen Bildgebung dadurch gekennzeichnet, dass die Merkmalsanzahl, also die Anzahl referenzierbarer Voxel, im sechsstelligen Bereich und höher liegt, während zumeist nur einige hundert Aufnahmen realisierbar sind. Selbst wenn nur eine sogenannte Region of Interest (ROI), also ein Teilgebiet des Gehirns, ausgewertet wird, übersteigt die Anzahl der Variablen oft die der Stichprobenelemente. Klassische multivariate Verfahren sind deshalb hierfür nicht anwendbar.

Standardmäßig erfolgt die Auswertung von fMRT-Daten daher voxelweise anhand eines univariaten allgemeinen linearen Modells, wobei die Korrelation der Stichprobenelemente – also die zeitliche Abhängigkeit – berücksichtigt wird. Hierfür wird in der Regel ein autoregressiver Prozess erster Ordnung (AR(1)-Prozess) unterstellt. Es hat sich etabliert, die Verletzung des Testniveaus bei der Durchführung von Homogenitätstests aufgrund der abhängigen Stichprobenelemente durch eine Satterthwaite-Approximation oder ein Pre-Whitening-Verfahren zu beheben (Friston *et al.*, 2007). Schließlich können die Testergebnisse für jedes Voxel visualisiert werden. Aufgrund der Vielzahl der Voxel geschieht dies in der Regel mit einem recht kleinen Schwellenwert ($p < 0,001$) oder unter multipler Betrachtung anhand von Adjustierungsmethoden für die familywise error rate (FWE) wie Bonferroni oder Gaussian Random Fields (Nichols & Hayasaka, 2003) oder die false discovery rate (FDR; Genovese *et al.*, 2001). Da aber auch die Voxel untereinander korrelieren, sind viele etablierte Adjustierungsverfahren, wie beispielsweise die Bonferroni-Korrektur, sehr konservativ. Daher schlagen unter anderem Chumbley *et al.* (2010) spezielle Adjustierungen für die funktionelle Bildgebung wie cluster- und voxelweise FDR-Methoden vor.

Bennett *et al.* (2009) kritisieren die jedoch noch immer praktizierte fMRT-Analyse anhand unadjustierter univariater Testverfahren. Durch den Nachweis neuronaler Aktivität bei einem toten Lachs machen die Autoren auf falsch-positive Resultate aufmerksam und fordern eine konsequente Adjustierung des multiplen Testens.

Nicht zuletzt aufgrund des multiplen Testproblems verstärken sich in jüngster Zeit die Bemühungen, multivariate Auswertungsmethoden einzusetzen, um räumliche beziehungsweise funktionelle Zusammenhänge zwischen den Voxeln zu berücksichtigen und somit eine höhere Güte zu erzielen. Beispielsweise konstruieren Kriegeskorte *et al.* (2006) ein sogenanntes Searchlight, mit dem Ziel anhand des Mahalanobis-Abstandes zweier Bedingungen für jedes Voxel die Kovarianzstruktur des es umgebenden Areal einzubeziehen. Weiterführend benutzen Hassabis *et al.* (2009) diesen Searchlight-Ansatz, um eine multivariate Pattern-Klassifikation zur Entdeckung neuronaler Netzwerke zu implementieren. Frackowiak *et al.* (2003) nutzen eine "eigenimage"-Analyse (eine Art Hauptkomponentenanalyse), um funktionell korrelierte Gebiete zu detektieren. Zum Teil sind

diese Verfahren aber dadurch beschränkt, dass die Anzahl der gleichzeitig betrachteten Voxel stets durch die Anzahl der vorhandenen Messungen begrenzt ist.

Weiterhin kommen Methoden wie Independent Component Analysis (ICA; [Frackowiak et al., 2003](#); [Sohr, 2007](#)) oder Bayes'sche Verfahren ([Penny et al., 2007](#)) zum Einsatz. Dies soll in dieser Arbeit jedoch nicht weiter verfolgt werden. Auch spätere Vergleiche mehrerer Probanden (sogenannte second-level-Analysen) sollen kein Bestandteil der Untersuchungen dieser Arbeit sein.

In der Diplomarbeit der Verfasserin dieser Arbeit ([Kose, 2007](#)) wird erstmals das Verhalten sogenannter stabilisierter multivariater Testverfahren von [Läuter et al. \(1996, 1998\)](#) bei zeitlich korrelierten Stichprobenelementen untersucht. Hierbei handelt es sich um Verfahren, die – im Falle unabhängiger Stichprobenelemente – bei hoher Güte auch dann das Testniveau exakt einhalten, wenn die Variablenanzahl deutlich über der Stichprobengröße liegt. Sowohl die Satterthwaite-Approximation als auch das Pre-Whitening-Verfahren werden in [Kose \(2007\)](#) in einem ersten Ansatz auf die stabilisierten multivariaten Verfahren angewendet und anhand von Simulationen dokumentiert.

Die vorliegende Arbeit verfolgt zwei Ziele: Einerseits eine wesentliche Erweiterung dieser ersten parametrischen Betrachtungen der stabilisierten multivariaten Verfahren bei abhängigen Stichprobenelementen. Dabei sollen statt der vereinfachten Anwendung des Zweistichprobenvergleichs Designs Beachtung finden, die die glättende Wirkung der zeitlichen Verzögerung in der neuronalen Reaktion auf verschiedene Reize berücksichtigen. Die bisher verwendeten Ansätze zur Schätzung der zeitlichen Korrelation werden diskutiert und erweitert. Zudem wird eine weitere Klasse multivariater Tests basierend auf paarweisen Abstandsmaßen in die Untersuchungen einbezogen, wobei erste Ergebnisse von [Naujoks \(2008\)](#) aufgegriffen und vertieft werden. Auch eine modifizierte Version der Principal Component Teststatistik ([Ding et al., 2012](#)), deren Verteilung insbesondere durch Rotationen realisiert werden kann, wird in diesem Zusammenhang behandelt.

Andererseits werden Permutationstests betrachtet, die als eine Alternative zur parametrischen Analyse von funktionellen MRT-Daten vorgeschlagen werden. Zur Analyse von Permutationstests im allgemeinen linearen Modell sind spezielle Überlegungen erforderlich. Die zeitliche Korrelation wird dabei zunächst als blockweise Permutation ([Kirch, 2007](#)) berücksichtigt. Die Einführung einer zusätzlichen zufälligen Verschiebung (Random Shift) optimiert diese Permutationsstrategie.

Alle parametrischen Methoden sowie der nichtparametrische Ansatz zur Analyse von funktionellen MRT-Daten werden sowohl an simulierten Daten als auch an realen Untersuchungsdaten gegenübergestellt.

Die Arbeit gliedert sich in folgende Abschnitte: Im Kapitel 2 werden die mathematischen Grundlagen beschrieben und die verwendeten Modelle und Testverfahren vorgestellt. Auf die Entstehung und Eigenschaften von fMRT-Messwerten, vor allem auf die sich dadurch ergebenden Modellbesonderheiten bei Analyse dieser Daten wird im Kapitel 3 eingegangen. Anschließend erfolgt in Kapitel 4 die Übersicht zu den in der univariaten Analyse von funktionellen Bilddaten etablierten parametrischen Adaptionsmethoden und deren Übertragung auf stabilisierte multivariate Testverfahren. Speziell die Verwendung und

Schätzung des zeitlichen Korrelationsmodells wird diskutiert und möglichen Alternativen gegenübergestellt. Kapitel 5 enthält die vorgeschlagene nichtparametrische Alternative im Umgang mit korrelierten Stichprobenelementen in Form einer Blockpermutation unter Verwendung eines Random Shift. Hier können vor allem Probleme bei der Schätzung der Korrelationsstruktur vermieden werden. Alle beschriebenen Verfahren werden in Kapitel 6 anhand simulierter Daten bezüglich der Einhaltung des Testniveaus (Fehler 1. Art) sowie der Güte (Fehler 2. Art) miteinander verglichen. Anschließend folgt die Anwendung der Methoden auf vorhandene Daten einer fMRT-Studie (Kapitel 7). Die Arbeit schließt in Kapitel 8 mit einer Zusammenfassung der Ergebnisse und einem kurzen Ausblick auf weiterführende Möglichkeiten der fMRT-Analyse.

2. Mathematische Grundlagen

Alle Überlegungen zur Testproblematik bei Daten der funktionellen Bildgebung in dieser Arbeit basieren auf dem multivariaten allgemeinen linearen Modell und entsprechenden Verteilungsannahmen. Daher werden in diesem Kapitel zunächst wichtige Aspekte zu den Verteilungen zusammengefasst. Danach wird das klassische multivariate lineare Modell betrachtet, gefolgt von einem Überblick über Testvorschläge für zweckmäßige Hypothesentests im Fall des klassischen allgemeinen linearen Modells. Auf die Betrachtung von Besonderheiten in der Modellbildung sowie bei den Hypothesentests für funktionelle MRT-Daten wird anschließend ausführlich in den Kapiteln 3, 4 und 5 eingegangen.

2.1. Verteilungen

In der Mehrzahl der in dieser Arbeit behandelten Ansätze wird von normalverteilten Daten ausgegangen. Dazu erfolgt zunächst die Charakterisierung der multi- wie auch matrixvariater Normalverteilung bei unabhängigen Stichprobenvektoren. Anschließend werden sphärische Verteilungen als Verallgemeinerung betrachtet, da sie die Grundlage der sogenannten stabilisierten Testverfahren darstellen, die bei der multivariaten Analyse der hochdimensionalen fMRT-Daten Anwendung finden.

2.1.1. Multi- und matrixvariater Normalverteilung

Im univariaten Fall ist die Normalverteilung durch die Dichtefunktion

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \quad \text{für } y \in \mathbb{R}$$

gekennzeichnet. Sie wird mit den Parametern $E(Y) = \mu$ für den Erwartungswert und $\text{Var}(Y) = \sigma^2$ für die Varianz anhand der Notation $Y \sim N(\mu, \sigma^2)$ charakterisiert (Fahrmeir et al., 2003).

Definition 2.1 Ein p -dimensionaler Vektor ist multivariat normalverteilt, wenn er die Dichtefunktion

$$f_{\mathbf{y}}(\mathbf{v}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{v} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{v} - \boldsymbol{\mu})\right\}, \quad \mathbf{v} \in \mathbb{R}^p$$

besitzt. Hierbei verwendet man die Notation $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, wobei $E(\mathbf{y}) = \boldsymbol{\mu}$ den p -dimensionalen Erwartungswertvektor und $\text{Var}(\mathbf{y}) = \boldsymbol{\Sigma}$ die Kovarianzmatrix vom Format $p \times p$ verkörpert (Handl, 2002). Dabei ist die Kovarianz positiv definit, d. h. $|\boldsymbol{\Sigma}| > \mathbf{0}$. Zudem bezeichnet $(\mathbf{v} - \boldsymbol{\mu})'$ den transponierten Vektor zu $\mathbf{v} - \boldsymbol{\mu}$.

Hierbei gilt:

Satz 2.1 *Ist \mathbf{y} ein p -variater normalverteilter Vektor mit $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, \mathbf{A} eine feste $k \times p$ -dimensionale Matrix vom Rang k und \mathbf{b} ein k -dimensionaler Vektor, dann ist $\mathbf{A}\mathbf{y} + \mathbf{b}$ k -variater normalverteilt mit $\mathbf{A}\mathbf{y} + \mathbf{b} \sim N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.*

Solange die betrachteten Stichprobenvektoren \mathbf{y}_i ($i = 1, \dots, n$) in einer Analyse unabhängig und identisch verteilt (iid) sind, ist deren Charakterisierung durch die multivariate Normalverteilung ausreichend. Da entsprechend dem Anwendungsgebiet der funktionellen Bildgebung jedoch in dieser Arbeit Situationen betrachtet werden, in denen die Erwartungswerte der einzelnen Stichprobenvektoren variieren können und auch die Unabhängigkeit der Stichprobenvektoren nicht mehr gewährleistet ist, wird hier der Begriff der matrixvariater Normalverteilung genutzt. Zu dessen Einführung sind zudem der sogenannte vec-Operator wie auch das Kronecker-Produkt notwendig (Fang & Zhang, 1990).

Definition 2.2 *Ist $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_p)$ eine Matrix vom Format $n \times p$, so ist der np -dimensionale Vektor $\text{vec}(\mathbf{A})$ durch*

$$\text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_p \end{bmatrix}$$

definiert, wobei vec als Operator zu verstehen ist.

Definition 2.3 *Sind $\mathbf{A} = (a_{ij})$ und \mathbf{B} Matrizen vom Format $n \times p$ beziehungsweise $m \times q$, so ist das Kronecker-Produkt dieser beiden Matrizen eine $nm \times pq$ -dimensionale Matrix mit*

$$\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B}) = \begin{bmatrix} a_{11}\mathbf{B} & \dots & a_{1p}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & \dots & a_{np}\mathbf{B} \end{bmatrix}.$$

Für eine reelle Zahl c gilt weiterhin

$$(c\mathbf{A}) \otimes \mathbf{B} = \mathbf{A} \otimes (c\mathbf{B}) = c(\mathbf{A} \otimes \mathbf{B}). \quad (2.1)$$

Mit Hilfe der Definitionen 2.2 und 2.3 lässt sich nun die matrixvariater Normalverteilung definieren (Fang & Zhang, 1990).

Definition 2.4 *Sei \mathbf{Y} eine $n \times p$ -dimensionale Zufallsmatrix. Falls $\text{vec}(\mathbf{Y}') \sim N_{np}(\boldsymbol{\mu}, \mathbf{A} \otimes \mathbf{B})$, wobei $\boldsymbol{\mu} = \text{vec}(\mathbf{M}')$, $\mathbf{M}: n \times p$, $\mathbf{A}: n \times n$, $\mathbf{B}: p \times p$ und \mathbf{A} wie auch \mathbf{B} positiv definit sind, hat \mathbf{Y} eine matrixvariater Normalverteilung und kann charakterisiert werden durch*

$$\mathbf{Y} \sim N_{n \times p}(\mathbf{M}, \mathbf{A} \otimes \mathbf{B}).$$

Dabei enthält die Matrix \mathbf{A} die Kovarianzstruktur der Zeilen und ist durch das Kronecker-Produkt verbunden mit der Matrix \mathbf{B} , der Kovarianzstruktur der Spalten.

Für die klassische Statistik sind die Zeilen als unabhängig vorausgesetzt, so dass $\mathbf{A} = \mathbf{I}_n$. Ist $\mathbf{Y} = (y_{ij}) \sim N_{n \times p}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{I}_p)$, dann sind alle einzelnen Elemente y_{ij} unabhängig und identisch verteilt nach $y_{ij} \sim N(0, 1)$.

Weiterhin gilt:

Satz 2.2 *Ist \mathbf{Y} matrixvariat normalverteilt nach $\mathbf{Y} \sim N_{n \times p}(\mathbf{M}, \mathbf{A} \otimes \mathbf{B})$ und sind \mathbf{C} , \mathbf{D} und \mathbf{F} feste Matrizen vom Format $m \times n$, $p \times q$ und $m \times q$, so ist $\mathbf{CYD} + \mathbf{F}$ verteilt nach*

$$\mathbf{CYD} + \mathbf{F} \sim N_{m \times q}(\mathbf{CMD} + \mathbf{F}, \mathbf{CAC}' \otimes \mathbf{D'BD}).$$

Dies bedeutet, dass eine Linksmultiplikation (\mathbf{CY}) sich in der Kovarianzstruktur des Produkts nur auf die Kovarianz zwischen den Stichprobenvektoren auswirkt, während eine Multiplikation von rechts (\mathbf{YD}) die Kovarianzstruktur zwischen den Variablen beeinflusst. Diese Eigenschaft der Normalverteilung wird später bei Erzeugung der Zufallszahlen genutzt (vgl. Abschnitt 6.1), indem zunächst iid standardnormalverteilte Zufallsvektoren simuliert und anschließend mit verschiedenen zeitlichen (zeilenspezifischen) wie auch räumlichen (spaltenspezifischen) Kovarianzstrukturen versehen werden. Des Weiteren basiert auch das im Abschnitt 4.1.2 vorgestellte Pre-Whitening, die Eliminierung von Korrelationen zwischen den Stichprobenvektoren, auf dieser Eigenschaft der Normalverteilung.

2.1.2. Linkssphärische Matrixverteilung

Bei der Durchführung von Tests mit hochdimensionalen Daten werden später auch Linearkombinationen analog zum Satz 2.2 genutzt, bei denen die Faktoren nun jedoch nicht mehr konstant sind, sondern von den Daten selbst abhängen. Diese Linearkombinationen sind nicht mehr normalverteilt. Unter bestimmten Bedingungen kann aber eine andere Klasse von Verteilungen herangezogen werden, die auch unter diesen Umständen noch exakte parametrische Tests ermöglicht: die linkssphärische Matrixverteilung (Läuter, 2005).

Definition 2.5 *Eine $n \times p$ -dimensionale Matrix \mathbf{Y} ($n \geq 1$, $p \geq 1$) heißt linkssphärisch verteilt, wenn sich durch jede Linksmultiplikation mit einer festen orthogonalen $n \times n$ -dimensionalen Matrix \mathbf{A} wieder die gleiche Verteilung ergibt: Es gilt*

$$\mathbf{AY} \stackrel{d}{=} \mathbf{Y} \tag{2.2}$$

für jede feste orthogonale $n \times n$ -dimensionale Matrix \mathbf{A} , also mit

$$\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}' = \mathbf{I}_n. \tag{2.3}$$

Ein Spezialfall einer linkssphärischen Verteilung ist eine Matrix aus n iid multivariat normalverteilten Vektoren mit Erwartung $\mathbf{0}$, die also entsprechend $\mathbf{Y} \sim N_{n \times p}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Sigma})$ verteilt ist. Denn diese Matrix ist bei jeder Linksmultiplikation mit einer festen orthogonalen $n \times n$ -dimensionalen Matrix \mathbf{A} wieder gleichermaßen matrixvariat normalverteilt mit Erwartung $\mathbf{0}$, da $E(\mathbf{AY}) = \mathbf{0}$ und $\text{Var}(\mathbf{AY}) = \mathbf{A}\mathbf{I}_n\mathbf{A}' \otimes \mathbf{\Sigma} = \mathbf{I}_n \otimes \mathbf{\Sigma}$ wegen (2.3).

2.2. Das klassische allgemeine lineare Modell

Funktionelle Bilddaten werden zumeist in Form des allgemeinen linearen Modells (GLM, general linear model) betrachtet. Dabei entsprechen die Messungen (Scans) den Stichprobenelementen, die Voxel den Variablen. Die Eigenschaften dieser Beobachtungen wie auch die der anderen Modellparameter weichen jedoch vom klassischen Modell ab, was letztlich die Intention dieser Arbeit zur Anpassung bestimmter Testverfahren an die funktionellen Bilddaten begründet.

Dementsprechend soll hier zunächst das klassische allgemeine lineare Modell vorgestellt werden, bevor in Kapitel 3 auf die Besonderheiten der Modelleigenschaften bei fMRT-Daten eingegangen wird.

Ein linearer Analyseansatz kann anhand des allgemeinen linearen Modells dargestellt werden. Dabei besteht die Beobachtungsmatrix \mathbf{Y} aus n Messungen über jeweils p Variablen. Sie wird erklärt durch eine Linearkombination verschiedener Einflussgrößen, die in der sogenannten Designmatrix enthalten sind, summiert mit einem zufälligen Fehler. Somit ergibt sich das im Folgenden definierte allgemeine lineare Modell.

Definition 2.6 *Das klassische allgemeine lineare Modell unter Annahme einer p -variablen Normalverteilung entspricht der Form*

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}, \quad \mathbf{E} \sim N_{n \times p}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Sigma}). \quad (2.4)$$

Dabei ist \mathbf{Y} die Matrix der Beobachtungen über n Messungen und p Variablen, \mathbf{X} die Designmatrix aus n Zeilen und s Spalten mit vollem Spaltenrang ($\text{Rg}(\mathbf{X}) = s$, $1 \leq s < n$) und \mathbf{B} die Parametermatrix vom Format $s \times p$. Weiterhin stellt \mathbf{E} die Matrix der Residuen mit n Zeilen und p Spalten dar. In deren Verteilung verkörpert \mathbf{I}_n die Kovarianzstruktur der Zeilen – im klassischen Fall deren Unabhängigkeit – und $\mathbf{\Sigma}$ die Kovarianzstruktur der Spalten.

Durch den vollen Spaltenrang der Designmatrix ist $\mathbf{X}'\mathbf{X}$ invertierbar. Die Unabhängigkeit der Spalten von \mathbf{X} kann dabei stets über eine entsprechende Parametrisierung des Problems erreicht werden. Dies wird in der fMRT-Anwendungssoftware wie SPM (SPM, 2009) aber nicht immer realisiert, um eine einfachere Interpretation der Ergebnisse zu bewirken. In solch einem Fall muss eine verallgemeinerte Inverse benutzt werden.

Für die Kovarianz ist hervorzuheben, dass die Kovarianzmatrix \mathbf{I}_n , die die Unabhängigkeit zwischen den n Stichprobenvektoren charakterisiert, durch das Kronecker-Produkt mit der Kovarianzstruktur der Spalten $\mathbf{\Sigma}$ verbunden ist. Dabei gilt analog zur Definition 2.3:

$$\mathbf{I}_n \otimes \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{\Sigma} \end{pmatrix}_{np \times np}. \quad (2.5)$$

In der klassischen Statistik sind die Stichprobenvektoren, wie im Modell (2.4) beschrieben, voneinander unabhängig. Dies ist Voraussetzung für die Durchführbarkeit der im

Abschnitt 2.3 vorgestellten klassischen Testverfahren, weshalb zunächst eine Beschränkung auf diese Unabhängigkeit der Messungen erfolgt.

Mit Hilfe des linearen Modells soll nun also geprüft werden, inwieweit die Beobachtungen durch die Spalten der Designmatrix, die bestimmten Einfluss- und Störgrößen entsprechen, beeinflusst werden. Dies geschieht über Hypothesentests zu Linearkombinationen der entsprechenden Parameter (Ahrens & Läuter, 1981).

Satz 2.3 *Bei Verwendung der Theorie der Kleinsten-Quadrate-Schätzer nach dem Prinzip der ordinary least squares (OLS), ergibt sich als Schätzung für die unbekannte Parametermatrix \mathbf{B} des allgemeinen linearen Modells nach (2.4),*

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \quad (2.6)$$

Ein Beweis dazu findet sich unter anderem in Fujikoshi *et al.* (2010).

Die zu testenden Hypothesen entsprechen Aussagen über spezielle Kontraste, das heißt Linearkombinationen der Spalten der Parametermatrix \mathbf{B}

$$H_0 : \mathbf{C}'\mathbf{B} = \mathbf{0}, \quad (2.7)$$

wobei \mathbf{C} die vorgegebene Kontrastgewichtsmatrix vom Format $s \times m$ ($s < m$) ist. Entsprechende klassische Teststatistiken zur Überprüfung solcher Hypothesen werden im Abschnitt 2.3 vorgestellt.

Mitunter werden die Einflussgrößen auch so definiert (Frackowiak *et al.*, 1997), dass auf die Verwendung der Kontrastgewichtsmatrix bei der Formulierung der Nullhypothese verzichtet werden kann. Dazu werden die Menge der Einflussgrößen und folglich die Designmatrix \mathbf{X} sowie die Parametermatrix \mathbf{B} in zwei unabhängige Teile zerlegt, so dass der zweite Teil nur solche Größen enthält, deren Einfluss unter der Nullhypothese verschwindet. Der erste Teil dagegen enthält jeweils nur Größen, die nicht in die Nullhypothese eingehen.

Mit $\mathbf{X}_{n \times s} = (\mathbf{X}_1 \mathbf{X}_2)$ und $\mathbf{B}_{s \times p} = \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix}$ ($\mathbf{X}_1: n \times s_1$, $\mathbf{X}_2: n \times s_2$, $\mathbf{B}_1: s_1 \times p$, $\mathbf{B}_2: s_2 \times p$, $s = s_1 + s_2$) entsteht dann das Modell

$$\mathbf{Y} = (\mathbf{X}_1 \mathbf{X}_2) \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix} + \mathbf{E} = \mathbf{X}_1 \mathbf{B}_1 + \mathbf{X}_2 \mathbf{B}_2 + \mathbf{E} \quad (2.8)$$

mit der interessierenden Nullhypothese

$$H_0 : \mathbf{B}_2 = \mathbf{0}_{s_2 \times p}. \quad (2.9)$$

In der Modellgleichung (2.8) verschwindet unter der Nullhypothese der Term $\mathbf{X}_2 \mathbf{B}_2$ und der Erwartungswert von \mathbf{Y} reduziert sich auf $E(\mathbf{Y}) = \mathbf{X}_1 \mathbf{B}_1$.

Wenn jedoch, um beispielsweise eine Eigenschaft einer linksphärisch verteilten Matrix zu erzeugen (vgl. Abschnitt 2.1.2), der Erwartungswert $\mathbf{0}$ benötigt wird, kann die verbleibende Parametermatrix \mathbf{B}_1 analog zu (2.6) durch $\hat{\mathbf{B}}_1 = (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{Y}$ geschätzt

werden. Damit kann dann die Beobachtungsmatrix \mathbf{Y} durch $\check{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}_1 \hat{\mathbf{B}}_1$ um diese Effekte bereinigt werden. Dies beschreibt die Multiplikation mit der entsprechenden residuenbildenden Matrix $\check{\mathbf{Y}} = \mathbf{R}_1 \mathbf{Y}$ mit $\mathbf{R}_1 = \mathbf{I}_n - \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1$ (vgl. Frackowiak *et al.*, 1997).

Beide Darstellungen des Testproblems lassen sich ineinander überführen. Das System (2.8)/(2.9) kann anhand der $s \times s_2$ -dimensionalen Kontrastgewichtsmatrix $\mathbf{C} = \begin{pmatrix} \mathbf{0}_{s_1 \times s_2} \\ \mathbf{I}_{s_2} \end{pmatrix}$ in die klassische Form (2.4)/(2.7) übertragen werden. Dabei gilt für die ursprüngliche Spaltendimension m von \mathbf{C} $m = s_2$.

Für die Umwandlung des klassischen (2.4)/(2.7) in das geteilte System (2.8)/(2.9) kann eine lineare Transformation der Einflussgrößen vorgenommen werden. Die für die Formulierung der Nullhypothese benötigten Linearkombinationen werden dabei von den restlichen Wirkungen der Faktoren separiert. Auch hier gilt $m = s_2$ und $s = s_1 + s_2$. Für die lineare Transformation wird die $s \times s_2$ -dimensionale Kontrastgewichtsmatrix \mathbf{C} durch $\mathbf{C}^* = \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}$ normiert und durch die $s \times s_1$ -dimensionale Matrix \mathbf{C}^{**} so ergänzt, dass $\mathbf{T}_{s \times s} = (\mathbf{C}^{**}\mathbf{C}^*)$ orthogonal ist.

Dann lässt sich die Modellgleichung (2.4) zu

$$\begin{aligned} \mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E} &= \mathbf{X}\mathbf{T}\mathbf{T}'\mathbf{B} + \mathbf{E} \\ &= \mathbf{X}(\mathbf{C}^{**}\mathbf{C}^*) \begin{pmatrix} \mathbf{C}^{**'} \\ \mathbf{C}^{*' } \end{pmatrix} \mathbf{B} + \mathbf{E} \\ &= (\check{\mathbf{X}}_1 \check{\mathbf{X}}_2) \begin{pmatrix} \check{\mathbf{B}}_1 \\ \check{\mathbf{B}}_2 \end{pmatrix} + \mathbf{E} \end{aligned}$$

umformulieren. Dabei entsprechen $(\check{\mathbf{X}}_{1_{n \times s_1}} \check{\mathbf{X}}_{2_{n \times s_2}})$ mit $\check{\mathbf{X}}_1 = \mathbf{X}\mathbf{C}^{**}$ und $\check{\mathbf{X}}_2 = \mathbf{X}\mathbf{C}^*$ sowie $\begin{pmatrix} \check{\mathbf{B}}_{1_{s_1 \times p}} \\ \check{\mathbf{B}}_{2_{s_2 \times p}} \end{pmatrix}$ mit $\check{\mathbf{B}}_1 = \mathbf{C}^{**'}\mathbf{B}$ und $\check{\mathbf{B}}_2 = \mathbf{C}^{*' }\mathbf{B}$ den transformierten Design- beziehungsweise Parameterteilmatrixen.

Die den ursprünglichen Kontrast betreffende Nullhypothese (2.7) ist äquivalent zu der nomierten $H_0 : \mathbf{C}^{*' }\mathbf{B} = \mathbf{0}$ und damit auch zu $H_0 : \check{\mathbf{B}}_2 = \mathbf{0}$, also der Form (2.9).

2.3. Testverfahren für das klassische allgemeine lineare Modell

Im vorherigen Kapitel werden das allgemeine lineare Modell vorgestellt und Hypothesen über die Parameterschätzer eingeführt. Entsprechende Hypothesentests für das klassische GLM (vgl. Abschnitt 2.2) sollen in diesem Kapitel kurz dargestellt werden. Dabei wird zunächst unterschieden zwischen univariaten Tests (Abschnitt 2.3.1), die nur eine Variable betrachten, und multivariaten Tests (Abschnitt 2.3.2). Bei den multivariaten Testverfahren wird zudem differenziert zwischen der klassischen Statistik, bei der mehr Stichprobenelemente als Variablen gefordert sind, und dem Fall, dass die Variablenanzahl die der Stichprobenelemente überschreitet. Da dies auf fMRT-Daten zutrifft, gilt

das Hauptaugenmerk dieser Arbeit den stabilisierten multivariaten Teststatistiken, die in diesem Fall weiterhin anwendbar sind. Zudem wird auf sogenannte Ähnlichkeits- und Distanzmaße eingegangen, die ebenfalls bei derart hochdimensionalen Daten angewandt werden können.

Anschließend soll eine Alternative zu den parametrischen Testverfahren in Form einer speziellen Permutation aufgezeigt werden, deren klassische Grundlagen im Abschnitt 2.3.3 dargestellt sind. Schließlich folgen Grundlagen von Rotationstests, die gewisse Ähnlichkeiten zum Permutationsprinzip aufweisen, jedoch auf parametrischen Annahmen basieren.

2.3.1. Univariate Analysen

Im univariaten Fall ($p = 1$), der bisher standardmäßig in der Anwendungssoftware wie SPM oder BrainVoyager zur Auswertung von fMRT-Daten implementiert ist (SPM, 2009; BrainInnovation, 2011), reduziert sich das Modell (2.4) bei unabhängigen Stichprobenelementen auf

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad (2.10)$$

wobei der Parametervektor analog zu (2.6) durch

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2.11)$$

geschätzt wird (Pruscha, 1996).

Die interessierende Nullhypothese zum univariaten Testproblem ist $H_0 : \mathbf{C}'\boldsymbol{\beta} = \mathbf{0}$ und kann je nach Dimension der Kontrastgewichtsmatrix $\mathbf{C}_{s \times m}$ anhand verschiedener Teststatistiken überprüft werden.

Im Fall $m = 1$ kann die klassische t -Teststatistik angewandt werden.

Satz 2.4 *Die Teststatistik*

$$t = \frac{\mathbf{c}'\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2 \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}} \quad (2.12)$$

mit der Varianzschätzung $\hat{\sigma}^2 = [(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})] / (n - s)$ und $\mathbf{c} \neq \mathbf{0}$ besitzt unter der Nullhypothese eine zentrale t -Verteilung mit $n - s$ Freiheitsgraden.

Den Beweis dazu liefert beispielsweise Fahrmeir *et al.* (2007).

Im allgemeinen Fall $m \geq 1$ verwendet man die F -Teststatistik, die die hypothesenbezogenen Abweichungen den Residuen gegenüberstellt.

Satz 2.5 *Die Teststatistik*

$$F = \frac{n - s}{m} \frac{h}{g} \quad (2.13)$$

folgt einer F -Verteilung mit m und $n - s$ Freiheitsgraden. Dabei entspricht h den Abweichungen von der Nullhypothese durch

$$\begin{aligned} h &= \hat{\boldsymbol{\beta}}' \mathbf{C} (\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C})^{-1} \mathbf{C}'\hat{\boldsymbol{\beta}} \\ &= \mathbf{y}'\mathbf{Q}_H\mathbf{y} \end{aligned} \quad (2.14)$$

mit

$$\mathbf{Q}_H = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C} \left(\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C} \right)^{-1} \mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (2.15)$$

und g den residualen Abweichungen durch

$$\begin{aligned} g &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}'\mathbf{Q}_G\mathbf{y} \end{aligned} \quad (2.16)$$

mit

$$\mathbf{Q}_G = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \quad (2.17)$$

Einen entsprechenden Beweis findet man in [Rasch \(1978\)](#).

Für $m = 1$ entspricht $F_{1,n-s} = t_{n-s}^2$ ([Sachs, 2004](#)).

2.3.2. Multivariate Analysen

Entsprechend dem multivariaten allgemeinen linearen Modell (2.4) werden n p -dimensionale Zeilenvektoren gleichzeitig betrachtet. Dabei sind zwei verschiedene Annahmen zu unterscheiden: klassische multivariate Verfahren können nur unter der Bedingung angewendet werden, dass die Anzahl p der Variablen die Differenz von Messungen und Parameterschätzern nicht übersteigt ($p \leq n - s$). Als Beispiel wird im Folgenden die Wilk's Λ -Teststatistik vorgestellt.

Übersteigt die Dimension der Variablen jedoch die Differenz $n - s$, arbeiten sogenannte stabilisierte multivariate Teststatistiken unter bestimmten Voraussetzungen dennoch exakt. Diese wie auch eine modifizierte Form der Tests werden betrachtet. Weiterhin sind unter diesen Umständen auch Tests auf der Basis von Ähnlichkeits- und Distanzmaßen anwendbar, die anschließend kurz beschrieben werden.

Klassische multivariate Analysen

Liegen n Stichprobenelemente vor, wobei $n - s \geq p$, können klassische multivariate Teststatistiken zum Testen der Nullhypothese verwendet werden. Eine weit verbreitete Statistik ist Wilks' Λ , in der die Determinanten der Produktsammenmatrizen \mathbf{G} und \mathbf{H} zueinander in Relation gesetzt werden ([Ahrens & Läuter, 1981](#); [Anderson, 1984](#)).

Satz 2.6 *Die Wilks' Λ -Statistik liefert einen exakten Test zum Prüfen der multivariaten Nullhypothese $H_0 : \mathbf{C}'\mathbf{B} = \mathbf{0}$. Sie ergibt sich aus*

$$\Lambda = \frac{|\mathbf{G}|}{|\mathbf{H} + \mathbf{G}|}. \quad (2.18)$$

Dabei sind \mathbf{H} und \mathbf{G} die multivariaten Pendanten zu (2.14) und (2.16), die durch

$$\mathbf{H} = \mathbf{Y}'\mathbf{Q}_H\mathbf{Y} \quad (2.19)$$

und

$$\mathbf{G} = \mathbf{Y}'\mathbf{Q}_G\mathbf{Y} \quad (2.20)$$

gebildet werden, wobei \mathbf{Q}_H und \mathbf{Q}_G den Gleichungen (2.15) und (2.17) entsprechen. Die Nullhypothese wird abgelehnt, wenn $\Lambda < U_{p,m,n-s}$, wobei $U_{p,m,n-s}$ das Quantil der Verteilung von Wilks' Λ ist (Anderson, 1984, Kapitel 8.4).

Für spezielle Fälle von Parameterwerten lässt sich Λ in eine F -verteilte Teststatistik umrechnen, beispielsweise:

$$\begin{aligned} p = 1 \quad m \text{ beliebig} & : \quad \frac{n-s}{m} \frac{1-\Lambda}{\Lambda} \sim F_{m,n-s} \quad , \\ p = 2 \quad m \text{ beliebig} & : \quad \frac{n-s-1}{m} \frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \sim F_{2m,2(n-s-1)} \quad , \\ p \text{ beliebig} \quad m = 1 & : \quad \frac{n-s-p+1}{p} \frac{1-\Lambda}{\Lambda} \sim F_{p,n-s-p+1} \quad , \\ p \text{ beliebig} \quad m = 2 & : \quad \frac{n-s-p+1}{p} \frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \sim F_{2p,2(n-s-p+1)} \quad . \end{aligned}$$

Andernfalls stehen gute F -Approximationen zur Verfügung (Pillai & Gupta, 1969; Ahrens & Läuter, 1981).

Neben Wilks' Λ gehören unter anderem das Spurkriterium von Pillai wie auch das von Hotelling zu den weit verbreiteten klassischen multivariaten Teststatistiken (Anderson, 1984), auf die hier aber nicht weiter eingegangen werden soll.

Stabilisierte multivariate Verfahren

Aufgrund der Dimension von fMRT-Daten – selbst bei Auswahl einer bestimmten Region of Interest – sind klassische multivariate Testverfahren nicht mehr anwendbar, da die Anzahl der Voxel die der Messungen weit übersteigt. Dies hätte am Beispiel von Wilks' Λ die Folge, dass die Produktsummenmatrizen \mathbf{H} und \mathbf{G} entsprechend (2.19) und (2.20) singularär würden.

Hier können aber weiterhin sogenannte stabilisierte multivariate Methoden angewendet werden, die von Läuter *et al.* (1996, 1998) vorgeschlagen werden. Diese Tests basieren auf univariaten oder niedrigdimensionalen (q , $1 \leq q \leq \min(p, n-s)$) linearen Scores (summarische Variablen), die mit Hilfe datenabhängiger Koeffizienten gebildet werden. Die Verfahren arbeiten auch bei hochdimensionalen Daten exakt, da den abgeleiteten Scores unter der Nullhypothese eine linkssphärische Verteilung (vgl. Abschnitt 2.1.2) zugrunde liegt, was für viele parametrische Tests ausreichend ist, um das Testniveau exakt einzuhalten. Die so gebildeten Scores fließen dann an Stelle von Originaldaten in klassische multivariate Teststatistiken wie zum Beispiel Wilks' Λ (2.18) ein (Läuter *et al.*, 1998).

Die $n \times q$ -dimensionalen Scores \mathbf{Z} für die Anwendung stabilisierter multivariater Tests werden durch Multiplikation der Matrix der Beobachtungen mit einer $p \times q$ -dimensionalen Gewichtsmatrix \mathbf{D} erzeugt

$$\mathbf{Z} = \mathbf{Y}\mathbf{D} . \quad (2.21)$$

Satz 2.7 *Hängt die Gewichtsmatrix \mathbf{D} von den Daten nur über die totale Produktsummenmatrix \mathbf{W} ab, wobei $\mathbf{W} = \mathbf{H} + \mathbf{G}$ mit \mathbf{H} und \mathbf{G} entsprechend (2.19) beziehungsweise (2.20) gebildet wird, und ist gewährleistet, dass \mathbf{Z} mit Wahrscheinlichkeit 1 den Rang q hat, dann halten die klassischen affin invarianten Tests (Wilks' Λ , Pillai-Spur, Hotelling-Spur) über die Scores anstelle der Beobachtungen das Testniveau exakt ein.*

Der Beweis dazu basiert auf der linkssphärischen Verteilung der Scores und ist beschrieben in [Läuter et al. \(1996, 1998\)](#).

Im einfachsten Fall kann ein eindimensionaler Score mit Hilfe eines Gewichtsvektors

$$\mathbf{d} = \text{Diag}(\mathbf{W})^{-\frac{1}{2}} \mathbf{1}_p$$

gebildet werden, wobei $\mathbf{1}_p$ ein p -dimensionaler Vektor bestehend aus Einsen ist. Dieser so erzeugte univariate Score kann nun beispielsweise anhand der t - (2.12) oder der F -Teststatistik (2.13) analysiert werden. Dieses Vorgehen beschreibt den sogenannten **Standardisierten Summentest (SS-Test)**, da die Effekte in den Daten nach einer Skalierung lediglich aufsummiert werden. Aus diesem Grund sollte der Test auch nur dann verwendet werden, wenn die Effekte in den verschiedenen Variablen gleichgerichtet und in ihrem Ausmaß (relativiert durch die jeweilige Standardabweichung) etwa gleich groß sind.

Eine alternative Teststatistik stellt der q -dimensionale **Principal Component Test (PC-Test)** dar. Hier wird die Gewichtsmatrix \mathbf{D} anhand des Eigenwertproblems der totalen Produktsummenmatrix \mathbf{W} gebildet ([Ahrens & Läuter, 1981](#)) und besteht aus den zu den q größten Eigenwerten korrespondierenden Eigenvektoren.

Definition 2.7 *Das Eigenwertproblem einer $p \times p$ -dimensionalen symmetrischen Matrix \mathbf{A} charakterisiert die Lösung der Gleichung $\mathbf{A}\mathbf{e}_i = \lambda_i\mathbf{e}_i$, $\mathbf{e}_i'\mathbf{e}_i = 1$, $i = 1, \dots, p$. Dabei sind λ_i die Eigenwerte und \mathbf{e}_i die dazugehörigen Eigenvektoren von \mathbf{A} . Mit $\boldsymbol{\lambda}' = (\lambda_1 \dots \lambda_p)$, $\boldsymbol{\Lambda} = \boldsymbol{\lambda}'\mathbf{I}_p$ und $\mathbf{E} = (\mathbf{e}_1 \dots \mathbf{e}_p)$ lässt sich das Eigenwertproblem auch schreiben als $\mathbf{A}\mathbf{E} = \mathbf{E}\boldsymbol{\Lambda}$ mit $\mathbf{E}'\mathbf{E} = \mathbf{I}_p$.*

Dabei sind die Eigenwerte in $\boldsymbol{\Lambda}$ absteigend sortiert: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Zudem gilt für zwei verschiedene Eigenwerte stets, dass ihre Eigenvektoren orthogonal zueinander sind, also $\mathbf{e}_i'\mathbf{e}_j = 0$ für $\lambda_i \neq \lambda_j$. Dadurch sind die Eigenvektoren im Fall vollends unterschiedlicher Eigenwerte bis auf das Vorzeichen eindeutig bestimmt ([Ahrens & Läuter, 1981](#)).

Vor allem bei hohen Variablenanzahlen kann zudem der sogenannte Dualismus von Eigenwertproblemen ausgenutzt werden ([Läuter, 2005](#)), um die Eigenvektoren der $p \times p$ -dimensionalen totalen Produktsummenmatrix \mathbf{W} aus dem dazugehörigen n -dimensionalen Eigenwertproblem zu bestimmen. Da

$$\begin{aligned} \mathbf{W} &= \mathbf{H} + \mathbf{G} \\ &= \mathbf{Y}'\mathbf{Q}_W\mathbf{Y} \\ &= (\mathbf{Q}_W\mathbf{Y})'\mathbf{Q}_W\mathbf{Y} \end{aligned}$$

geschrieben werden kann und die Menge q der positiven Eigenwerte für $(\mathbf{Q}_W \mathbf{Y})' \mathbf{Q}_W \mathbf{Y}$ und $\mathbf{Q}_W \mathbf{Y} (\mathbf{Q}_W \mathbf{Y})'$ übereinstimmt, können die zusammengehörigen Eigenwertprobleme

$$((\mathbf{Q}_W \mathbf{Y})' \mathbf{Q}_W \mathbf{Y}) \mathbf{D}_{p \times q} = \mathbf{D}_{p \times q} \mathbf{\Lambda}, \quad \mathbf{D}'_{p \times q} \mathbf{D}_{p \times q} = \mathbf{I}_q$$

und

$$(\mathbf{Q}_W \mathbf{Y} (\mathbf{Q}_W \mathbf{Y})') \mathbf{D}_{n \times q} = \mathbf{D}_{n \times q} \mathbf{\Lambda}, \quad \mathbf{D}'_{n \times q} \mathbf{D}_{n \times q} = \mathbf{I}_q$$

ineinander transformiert werden. Der q -dimensionale Eigenvektor des Ausgangsproblems kann somit über das kleinere Eigenwertproblem

$$\mathbf{D}_{p \times q} = (\mathbf{Q}_W \mathbf{Y})' \mathbf{D}_{n \times q} \begin{pmatrix} \mathbf{\Lambda}_{q \times n}^{-1/2} & \mathbf{0}_{q \times (q-n)} \end{pmatrix}$$

berechnet werden, wodurch sich für $n \ll p$ der rechentechnische Aufwand stark vermindert.

Beim der skaleninvarianten Version des PC-Tests ist \mathbf{D} Lösung des verallgemeinerten Eigenwertproblems $\mathbf{W} \mathbf{D} = \text{Diag}(\mathbf{W}) \mathbf{D} \mathbf{\Lambda}$ und erfüllt $\mathbf{D}' \text{Diag}(\mathbf{W}) \mathbf{D} = \mathbf{I}_q$, wobei $\mathbf{\Lambda}$ die q größten Eigenwerte von \mathbf{W} in der Diagonalen enthält und \mathbf{D} selbst aus den dazugehörigen Eigenvektoren besteht. Analog zur skalenabhängigen Variante kann der Dualismus des verallgemeinerten Eigenwertproblems für eine effektivere Programmierung ausgenutzt werden (Läuter, 2005).

Bei Verwendung dieser Version des PC-Tests werden unterschiedliche Skalen der Variablen berücksichtigt. Bei etwa gleichen Skalenniveaus hingegen kann eine skalenabhängige Version des PC-Tests benutzt werden, bei der die Gewichtsmatrix über das einfache Eigenwertproblem $\mathbf{W} \mathbf{D} = \mathbf{D} \mathbf{\Lambda}$ mit $\mathbf{D}' \mathbf{D} = \mathbf{I}_q$ gewonnen wird. Für die in dieser Arbeit untersuchten fMRT-Daten ist diese Variante des PC-Tests voraussichtlich zweckmäßig, da die gemessenen BOLD-Signale innerhalb des gesamten Gehirns einer einheitlichen Skala folgen.

Bei beiden Varianten des PC-Tests werden die Scores anhand der Gewichtsmatrix entsprechend (2.21) berechnet. Die Dimension q der Scores wird zuvor über ein gängiges Kriterium wie das von Kaiser (1960) oder Jolliffe (1972) in Abhängigkeit von \mathbf{W} festgelegt, dabei gilt $q \leq \min(p, n - s)$. Anschließend kann der Score über eine klassische multivariate Statistik wie Wilks' Λ (2.18) getestet werden, wobei der Parameter der Dimension p durch q ersetzt wird.

Die Auswahl der Dimension q anhand des Kriteriums von Kaiser (1960) entspricht der Anzahl der Eigenwerte der totalen Produktsummenmatrix \mathbf{W} , die größer sind als der Mittelwert aller Eigenwerte, also $q = \max(i \mid \lambda_i > \bar{\lambda}), i = 1, \dots, p$, mit $\bar{\lambda} = \frac{1}{p} \sum_{i=1}^p \lambda_i$.

Jolliffe (1972) zeigt anhand von Simulationsstudien für Wetterdaten, dass dieses Kriterium zu hart sein kann und schlägt eine Abwandlung des Kaiser-Kriteriums in Form von $q = \max(i \mid \lambda_i > 0, 7\bar{\lambda}), i = 1, \dots, p$, vor, demzufolge mehr Variablen eingeschlossen werden können.

Zaino (2006) beschreibt indes, dass gerade bei medizinischen Daten schon das Kaiser-Kriterium eine recht hohe Scoredimension liefert und schlägt daher die Verwendung eines

inversen Jolliffe-Kriteriums

$$q = \max(i \mid \lambda_i > \frac{\bar{\lambda}}{0,7}), i = 1, \dots, p \quad (2.22)$$

vor, das im Weiteren in dieser Arbeit zur Bestimmung der Dimension des PC-Tests verwendet wird. Für hochdimensionale Daten ($p > n - s$) sind jedoch die letzten $p - (n - s + 1)$ Eigenwerte Null, wodurch sich der Mittelwert über alle Eigenwerte bei einer hohen Variablenanzahl stark verringern würde. Daher ergibt sich $\bar{\lambda}$ in diesem Fall nur aus den Eigenwerten ungleich Null.

Am Beispiel von Wilks' Λ kann dargestellt werden, dass die Scorebildung unmittelbar in die Teststatistik eingearbeitet werden kann. Ein separater vorgelagerter Schritt der Datenbehandlung ist somit nicht erforderlich. Entsprechend (2.19) und (2.20) gilt mit (2.21) für die auf die Scores angewendeten Quadratsummen

$$\mathbf{H}_Z = \mathbf{Z}'\mathbf{Q}_H\mathbf{Z} = \mathbf{D}'\mathbf{Y}'\mathbf{H}\mathbf{Y}\mathbf{D} = \mathbf{D}'\mathbf{H}\mathbf{D} \quad (2.23)$$

und

$$\mathbf{G}_Z = \mathbf{Z}'\mathbf{Q}_G\mathbf{Z} = \mathbf{D}'\mathbf{Y}'\mathbf{G}\mathbf{Y}\mathbf{D} = \mathbf{D}'\mathbf{G}\mathbf{D}, \quad (2.24)$$

so dass Wilks' Λ basierend auf den Scores als

$$\Lambda = \frac{|\mathbf{D}'\mathbf{G}\mathbf{D}|}{|\mathbf{D}'\mathbf{H}\mathbf{D} + \mathbf{D}'\mathbf{G}\mathbf{D}|} \quad (2.25)$$

geschrieben werden kann, wobei die Dimension p in den zuvor dargestellten Verteilungsbetrachtungen durch q ersetzt wird.

Modifizierte Version des PC-Tests

Zusätzlich kommt eine modifizierte Version des q -dimensionalen Principal Component Tests zum Einsatz, die von Ding *et al.* (2012) vorgeschlagen wird. Hierbei wird eine gewichtete Kombination der Quadratsummen für die unterschiedliche Hauptkomponenten genutzt, um so bei geringer Stichprobenanzahl eine höhere Güte zu erzielen. Dafür wird nach der Reduktion der Daten auf einen q -dimensionalen Score mit $q > 1$ eine modifizierte univariate F -Statistik analog zu (2.13) nutzbar gemacht, wobei die Verteilungseigenschaften dann angepasst werden müssen.

Eine genauere Erklärung dessen erfolgt wiederum auf Grundlage des geteilten Modells (2.8). Bei Reduktion der $n \times p$ -dimensionalen Daten auf einen eindimensionalen Score \mathbf{z} wäre beispielsweise die entsprechend (2.13) gebildete F -Statistik

$$F = \frac{n - s_1 - s_2}{s_2} \frac{h_z}{g_z} \sim F_{s_2, n - s_1 - s_2} \quad (2.26)$$

verwendbar, wobei h_z und g_z die univariaten Pendanten zu (2.23) und (2.24) sind. Im multivariaten Fall werden nun in (2.26) die Skalare h_z und g_z durch eine gewichtete Summe der Diagonalelemente von (2.23) und (2.24) ersetzt. Die Gewichtung erfolgt

anhand der entsprechenden q größten Eigenwerte λ_i , $i = 1, \dots, q$, der totalen Produktsummenmatrix \mathbf{W} . Damit entsteht die Teststatistik

$$\tilde{F} = \frac{n - s_1 - s_2}{s_2} \frac{\sum_{i=1}^q \lambda_i \mathbf{H}_{\mathbf{Z}_{i,i}}}{\sum_{i=1}^q \lambda_i \mathbf{G}_{\mathbf{Z}_{i,i}}}. \quad (2.27)$$

Diese Statistik folgt im Allgemeinen keiner F -Verteilung mehr. Man kann aber eine Satterthwaite-Approximation einer F -Verteilung mit vs_2 und $v(n - s_1 - s_2)$ Freiheitsgraden angeben, wobei

$$v = \frac{\left(\sum_{i=1}^q \lambda_i \right)^2}{\sum_{i=1}^q \lambda_i^2}$$

(Ding *et al.*, 2012). Um unter Verwendung dieser Teststatistik auch einen exakten Test zu erzielen, schlagen Ding *et al.* (2012) vor allem bei sehr geringem Stichprobenumfang die Verwendung eines Rotationstests (vgl. Abschnitt 2.3.4 der vorliegenden Arbeit) vor. Alternativ könnte bei ausreichendem Stichprobenumfang ebenso ein Permutationsverfahren (vgl. Abschnitt 2.3.3) genutzt werden.

Ähnlichkeits- und Distanzmaße

Um Gruppenunterschiede zwischen p -dimensionalen Vektoren aufzuzeigen, können auch Teststatistiken benutzt werden, die auf sogenannte Ähnlichkeits- oder Distanzmaße (Hartung, 1986; Handl, 2002; Naujoks, 2008) aufbauen. Der Grundgedanke hierbei ist, dass sich Vektoren gleicher Gruppen zueinander ähnlicher verhalten als Vektoren unterschiedlicher Gruppen. Im Fall von funktionellen MRT-Daten könnten somit beispielsweise Reiz- und Ruhebedingungen gegeneinander getestet werden. Die Länge p der Vektoren ist dabei nicht durch die Stichprobenanzahl n begrenzt, da das Maß für die Ähnlichkeit oder Distanz zweier Stichprobenvektoren unabhängig von deren Länge immer ein Skalar ist. Dabei ist auch eine Übertragung dieses Vorgehens auf Tests im allgemeinen linearen Modell möglich (Kropf & Adolf, 2009).

Ein Ähnlichkeitsmaß $s_{ij} = s(\mathbf{y}_i, \mathbf{y}_j)$ misst die Ähnlichkeit zweier Stichprobenvektoren \mathbf{y}_i und \mathbf{y}_j , $i, j = 1, \dots, n$. Je ähnlicher sich die Stichprobenvektoren sind, desto größer wird das Maß. Gleiche Stichprobenvektoren weisen die größtmögliche Ähnlichkeit auf. Ähnlichkeitsmaße erfüllen daher in der Regel die folgenden Eigenschaften:

- $s_{ii} = 1$ ($i = 1, \dots, n$),
- $s_{ii} \geq s_{ij}$ ($i, j = 1, \dots, n$) sowie
- $s_{ij} = s_{ji}$ ($i, j = 1, \dots, n$).

Beispielgebend für die Ähnlichkeitsmaße wird in dieser Arbeit der Korrelationskoeffizient r_{ij} nach Pearson betrachtet (Handl, 2002). Zwischen den beiden p -dimensionalen Stichprobenvektoren \mathbf{y}_i und \mathbf{y}_j , $i, j = 1, \dots, n$, berechnet er sich anhand der Gleichung

$$r_{ij} = \frac{\sum_{k=1}^p (y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j)}{\sqrt{\left(\sum_{l=1}^n (y_{il} - \bar{y}_i)^2\right) \left(\sum_{l=1}^n (y_{jl} - \bar{y}_j)^2\right)}}. \quad (2.28)$$

Dabei ist $\bar{y}_i = \frac{1}{p} \sum_{l=1}^p y_{il}$ das arithmetische Mittel des i -ten Stichprobenvektors ($i = 1, \dots, n$).

Im Gegensatz dazu ermittelt ein Distanzmaß $t_{ij} = t(\mathbf{y}_i, \mathbf{y}_j)$ die Distanz, also die Unähnlichkeit, zweier Stichprobenvektoren \mathbf{y}_i und \mathbf{y}_j , $i, j = 1, \dots, n$. Damit wird es umso größer, je unähnlicher sich die Stichprobenvektoren sind. Gleiche Vektoren weisen dementsprechend die kleinstmögliche Distanz auf, somit ergeben sich für Distanzmaße die Eigenschaften:

- $t_{ii} = 0$ ($i = 1, \dots, n$),
- $t_{ii} \leq t_{ij}$ ($i, j = 1, \dots, n$) sowie
- $t_{ij} = t_{ji}$ ($i, j = 1, \dots, n$).

Distanzmaße sind häufig Abstände in \mathbb{R}^p , wie beispielsweise der quadratische euklidische Abstand q_{ij} (Handl, 2002), der in dieser Arbeit betrachtet wird. Geometrisch gesehen ($p = 1, 2$ oder 3) entspricht er dem Quadrat der kleinsten Differenz der beiden Stichprobenvektoren \mathbf{y}_i und \mathbf{y}_j , $i, j = 1, \dots, n$, in \mathbb{R}^p und wird entsprechend

$$q_{ij} = \sum_{k=1}^p (y_{ik} - y_{jk})^2 \quad (2.29)$$

berechnet.

Die Ähnlichkeits- oder Distanzmaße zwischen allen n Stichprobenvektoren werden in einer sogenannten Ähnlichkeitsmatrix \mathbf{S} beziehungsweise der Distanzmatrix \mathbf{T} dargestellt

$$\mathbf{S} = \begin{pmatrix} s_{11} & \dots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \dots & s_{nn} \end{pmatrix}, \quad \mathbf{T} = \begin{pmatrix} t_{11} & \dots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \dots & t_{nn} \end{pmatrix}.$$

Auf deren Basis kann nun eine Teststatistik kreiert werden, die die Abstände von Vektoren innerhalb derselben Gruppe mit den Abständen von Vektoren aus verschiedenen Gruppen vergleicht (Anderson, 2001).

In Kropf *et al.* (2004, 2007) wird beispielsweise die Differenz des mittleren Maßes innerhalb der Gruppen mit dem mittleren Maß zwischen verschiedenen Gruppen betrachtet

$$d = \bar{r}_{within} - \bar{r}_{between}. \quad (2.30)$$

Bei diesem Beispiel wäre \bar{r}_{within} das arithmetische Mittel aller Korrelationskoeffizienten der Paare von Stichprobenvektoren aus ein- und derselben Gruppe (wobei die Einselemente auf der Diagonalen für identische Paare ausgelassen werden) und $\bar{r}_{between}$ analog dazu das Mittel der Vektorenpaare aus unterschiedlichen Gruppen.

Anschließend kann für diese Teststatistik ein Permutationstest (vgl. Abschnitt 2.3.3) oder ein Rotationstest (vgl. Abschnitt 2.3.4) genutzt werden, um unter der Nullhypothese der Gleichheit der Gruppen einen P -Wert zu ermitteln.

In den Simulationen dieser Arbeit wird die Korrelation der Ähnlichkeits- beziehungsweise Distanzmaße für Paare von Stichprobenvektoren mit den absoluten Differenzen im Zielkontrast der Designmatrix gebildet und über einen Permutationstest geprüft.

2.3.3. Permutationstests

Neben den zuvor vorgestellten parametrischen Teststatistiken, die auf bestimmten Verteilungsannahmen der Daten basieren, werden in dieser Arbeit auch nichtparametrische Tests in Form von Permutationen einbezogen, bei denen diese Annahmen durch schwächere ersetzt werden. Dabei sind innerhalb eines Permutationstests sehr flexible gestaltete Teststatistiken nutzbar.

Grundlagen der Permutation

Im Gegensatz zu parametrischen Tests basieren Permutationsstrategien nicht auf der Unabhängigkeit der Beobachtungen und dies entsprechend einer bestimmten Verteilungsannahme, sondern setzen nur die Austauschbarkeit aufeinanderfolgender Beobachtungen unter der Nullhypothese voraus (Bernardo, 1996; Good, 2005).

Im Falle multivariater Daten entsprechen die Beobachtungen p -dimensionalen Zeilenvektoren. Diese werden jeweils als eine Einheit betrachtet (Good, 2005), so dass sich eine Permutation auf alle Elemente des Vektors einheitlich auswirkt.

Definition 2.8 *Beobachtungsvektoren $\mathbf{y}'_1, \dots, \mathbf{y}'_n$ sind austauschbar, wenn ihre gemeinsame Dichtefunktion $f(\mathbf{y}'_1, \dots, \mathbf{y}'_n)$ unabhängig von der Ordnung, also für alle Permutationen π eines Sets der Reihenfolgen $\{1, \dots, n\}$ gleich ist:*

$$f(\mathbf{y}'_1, \dots, \mathbf{y}'_n) = f(\mathbf{y}'_{\pi(1)}, \dots, \mathbf{y}'_{\pi(n)}).$$

Permutationsstrategien eignen sich beispielsweise gut für Homogenitätstests bei unabhängigen Teilstichproben. Dabei wird als Nullhypothese die Austauschbarkeit der Beobachtungen auch zwischen verschiedenen Teilstichproben unterstellt. Unter den sonst üblichen Annahmen unabhängiger Stichprobenelemente und gleicher Verteilungen in den verschiedenen Stichproben unter der entsprechenden Nullhypothese ist dies gewährleistet. Nun wird eine geeignete Teststatistik sowie deren Verteilung unter der Nullhypothese bestimmt.

Für die weiteren Betrachtungen wird dabei unterstellt, dass die Teststatistik so bestimmt ist, dass für sie bei Abweichungen von der Nullhypothese größere Werte zu erwarten sind. Dies lässt sich immer durch eine geeignete Festlegung der Testgröße erzielen.

Die Permutationsverteilung erhält man, indem die Teststatistik für alle Permutationen

der Stichprobe ermittelt wird. Die Permutationsverteilung entspricht damit der bedingten Verteilung der Teststatistik bei gegebenen Werten (aber zufälligen Permutationen) der Stichprobenelemente beziehungsweise -vektoren.

Entspricht die Teststatistik basierend auf den Originalwerten mindestens dem Wert des $(1 - \alpha)$ -Quantils der Permutationsverteilung, kann die Nullhypothese austauschbarer Beobachtungen abgelehnt werden (Basso *et al.*, 2009).

Entsprechende Angaben zur Durchführung eines Permutationstest finden sich bei Good (2005), der explizite Schritte für einen Permutationstest beschreibt:

1. Festlegung von Null- und Alternativhypothese
2. Auswahl einer geeigneten Teststatistik
3. Berechnung der Teststatistik für die Originalwerte
4. Permutation der Originalwerte
Berechnung der Teststatistik für die permutierten Beobachtungen
Vergleich dieser Teststatistik mit der basierend auf den Originalwerten
Wiederholung dieses Schrittes für alle Permutationen
5. Entscheidung bezüglich der Nullhypothese

Die Nullhypothese kann abgelehnt werden, wenn anhand der Originalwerte ein im Rahmen der Permutationsverteilung extremer Wert der Teststatistik entsteht. Dabei wird gezählt, wie oft die Teststatistik aus allen permutierten Beobachtungen T_π einschließlich der Originalstichprobe selbst größer oder gleich T_0 ist. Dabei entspricht T_0 der Statistik basierend auf den Originalwerten. Das Verhältnis dieser Summe zur Gesamtanzahl aller möglichen Permutationen (im Weiteren mit N_π bezeichnet) ergibt dann den P -Wert:

$$P = (\#(T_\pi \geq T_0)) / N_\pi \quad (2.31)$$

Signifikanz ($P \leq \alpha$) erreicht man somit in Übereinstimmung mit der obigen Angabe, wenn die originale Teststatistik mindestens dem $(1 - \alpha)$ -Quantil der Permutationsverteilung entspricht. Ob mit einem P -Wert nach (2.31) jedoch überhaupt Signifikanz erreicht werden kann, hängt direkt von der Anzahl der möglichen Permutationen und somit vom Stichprobenumfang ab. Bei beispielsweise nur sechs Permutationen kann nach (2.31) minimal ein P -Wert von $0,1\bar{6}$ erreicht werden. Ein signifikantes Ergebnis ist demnach nur bei einem hinreichend großen Stichprobenumfang überhaupt erzielbar.

Die Anzahl möglicher Permutationen ergibt sich aus der Anzahl der zu permutierenden Beobachtungen (Fahrmeir *et al.*, 2003).

Satz 2.8 Besteht die zu permutierende Stichprobe $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ aus n Elementen, gibt es genau $n!$ Möglichkeiten der Permutation.

Satz 2.8 gilt, da die Permutation eines Stichprobenvektors der Länge n dem Fall der Auswahl von n Elementen ohne Zurücklegen entspricht: Alle Elemente werden permutiert, wobei jedes Element nur einmal ausgewählt werden kann. Folglich ergeben sich für die erste Auswahl n Möglichkeiten, anschließend nur noch $(n - 1)$, $(n - 2)$, usw. und damit eine Gesamtpermutationsanzahl von $n! = n(n - 1)(n - 2) \dots 1$.

Für beispielsweise nur zehn Elemente sind somit bereits 3 628 800 verschiedene Permutationen möglich.

Im Fall der Anwendung von Permutationen für Homogenitätstests von unabhängigen Teilstichproben, also dem Vergleich der Mittelwerte der Teilstichproben, ist zu beachten, dass eine Permutation ausschließlich innerhalb der einzelnen Teilstichproben keine Veränderung der Teststatistik bewirkt. Damit reduziert sich die Anzahl effektiver Permutationen eines Stichprobenvektors der Länge n ($n = n_1 + n_2 + \dots + n_k$) bei k Teilstichproben von $n!$ auf

$$\frac{n!}{\prod_{i=1}^k n_i!}$$

(Mielke & Berry, 2007). Für den Spezialfall von $k = 2$ Teilstichproben entspricht dies

$$\frac{(n_1 + n_2)!}{n_1!n_2!} = \binom{n_1 + n_2}{n_1} = \binom{n_1 + n_2}{n_2}.$$

Abwandlungen des Permutationsprinzips

a) Monte-Carlo-Simulation

Selbst bei einer relativ kleinen Anzahl von Stichprobenelementen ist die Anzahl aller möglichen Permutationen (N_π) enorm hoch. Damit steigt die benötigte Rechenzeit bei Durchlauf aller Permutationen in teilweise unrealisierbare Höhe. Um dies zu umgehen, kommt oftmals eine sogenannte **Monte-Carlo-Simulation** zum Einsatz (Good, 2005). Dabei wird eine ausreichend große Anzahl zufälliger Permutationen (n_π) simuliert, die repräsentativ für die komplette Permutationsverteilung herangezogen wird. Die Anzahl zufälliger Simulationen wird in der Regel zuvor festgelegt und sollte generell nicht unter $n_\pi = 1\,000$ liegen (Anderson, 2001).

Die Teststatistik aus der Originalstichprobe T_0 wird gemeinsam mit den n_π Werten von Teststatistiken T_π aus den permutierten Daten in die Berechnung des P -Wertes einbezogen

$$P_{MC} = \frac{(\#(T_\pi \geq T_0)) + 1}{n_\pi + 1}, \quad (2.32)$$

da diese $n_\pi + 1$ Werte unter der Nullhypothese die gleiche Verteilung besitzen.

Bei dieser Berechnung des P -Wertes erhält man immer einen exakten Test, beziehungsweise einen leicht konservativen, da man im Allgemeinen nicht kontrolliert, ob ein- und dieselbe zufällige Simulation mehrfach auftaucht (Phipson & Smyth, 2010).

Die Abweichung von (2.31), dem P -Wert basierend auf allen möglichen Permutationen, lässt sich durch eine Binomialverteilung der Größe $\#(T_\pi \geq T_0)$ beschreiben. Hieraus

folgt die Standardabweichung

$$\sigma_P = \sqrt{\frac{P_{MC}(1 - P_{MC})}{n_\pi}} \quad (2.33)$$

für die Differenz von P_{MC} aus (2.32) zum "wahren" P -Wert nach (2.31).

b) Permutationen innerhalb der Designmatrix

Für die folgenden Betrachtungen des Permutationsprinzips soll wieder auf das in dieser Arbeit grundlegende Modell (2.4)

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}, \quad \mathbf{E} \sim N_{n \times p}(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma})$$

zurückgegriffen werden. Allerdings wird hier zunächst die globale Nullhypothese

$$H_0 : \mathbf{B} = \mathbf{0} \quad (2.34)$$

betrachtet, also dass keine der Kovariablen einen Einfluss hat.

Die für die Permutation notwendige Annahme der Austauschbarkeit der Stichprobenvektoren ist hierbei gewährleistet, da unter der Nullhypothese alle Stichprobenvektoren unabhängig identisch normalverteilt sind mit

- $E(\mathbf{y}_i) = \mathbf{0} \quad (i = 1, \dots, n)$
- $\text{Var}(\mathbf{y}_i) = \boldsymbol{\Sigma} \quad (i = 1, \dots, n)$
- $\text{Cov}(\mathbf{y}_i, \mathbf{y}_j) = \mathbf{0} \quad (i, j = 1, \dots, n; i \neq j) \quad .$

Durch eine Permutation der Stichprobenvektoren in diesem Modell ändert sich die Zuordnung der Beobachtungen zu ihren erklärenden Anteilen aus dem linearen Modell. Für den i -ten Vektor der Originaldaten ergibt sich

$$\mathbf{y}_i = \mathbf{x}_i\mathbf{B} + \boldsymbol{\epsilon}_i, \quad (2.35)$$

für den Vektor an derselben Stelle bei Permutation demnach

$$\mathbf{y}_{\pi(i)} = \mathbf{x}_i\mathbf{B} + \boldsymbol{\epsilon}_{\pi(i)}. \quad (2.36)$$

Die permutierte Beobachtungsmatrix $\mathbf{Y}_\pi = \begin{pmatrix} \mathbf{y}_{\pi(1)} \\ \vdots \\ \mathbf{y}_{\pi(n)} \end{pmatrix}$ besitzt die gleichen Verteilungseigenschaften wie die originalen Beobachtungswerte, es gilt $\mathbf{Y} \stackrel{d}{=} \mathbf{Y}_\pi$. Weiterhin gilt für jede Teststatistik als Funktion der Beobachtungswerte und der Designmatrix \mathbf{X}

$$T(\mathbf{Y}, \mathbf{X}) \stackrel{d}{=} T(\mathbf{Y}_\pi, \mathbf{X}). \quad (2.37)$$

Da die im linearen Modell verwendeten Teststatistiken unabhängig von der Anordnung der Zeilen in \mathbf{X} und \mathbf{Y} sind, sofern diese nur simultan verändert werden, kann man diese Reihenfolge durch die Anwendung der zu π inversen Permutation π^{-1} wieder so ändern, dass in \mathbf{Y} die Zeilen in der ursprünglichen Reihenfolge stehen, dafür aber \mathbf{X} invers permutiert ist

$$T(\mathbf{Y}_\pi, \mathbf{X}) = T(\mathbf{Y}, \mathbf{X}_{\pi^{-1}}). \quad (2.38)$$

Auch die inverse Permutation ist eine Permutation, so dass die Verteilungseigenschaften unter der Nullhypothese beibehalten werden. Daher kann bei Verwendung der globalen Nullhypothese (2.34) anstelle einer Permutation der Beobachtungsvektoren auch eine **Permutation der Zeilen der Designmatrix** erfolgen, um die Permutationsverteilung abzuleiten.

Zur Untersuchung der allgemeinen Nullhypothese $H_0 : \mathbf{C}'\mathbf{B} = \mathbf{0}$ für das Modell mit Hilfe von Permutationen wird die im Abschnitt 2.2 beschriebene lineare Transformation durchgeführt, um das mittels der Kontrastgewichtsmatrix \mathbf{C} formulierte Problem auf das geteilte Modell (2.8)

$$\mathbf{Y} = (\mathbf{X}_1 \mathbf{X}_2) \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix} + \mathbf{E}$$

mit der korrespondierenden Nullhypothese (2.9) $H_0 : \mathbf{B}_2 = \mathbf{0}$ zurückzuführen. Im Ergebnis entfällt unter der Nullhypothese der Anteil $\mathbf{X}_2 \mathbf{B}_2$, der Anteil $\mathbf{X}_1 \mathbf{B}_1$ bleibt jedoch erhalten, so dass $E(\mathbf{Y}) = \mathbf{X}_1 \mathbf{B}_1$.

Es ergibt sich statt (2.35) für den i -ten Vektor der Originaldaten

$$\mathbf{y}_i = \mathbf{x}_{1_i} \mathbf{B}_1 + \mathbf{x}_{2_i} \mathbf{B}_2 + \boldsymbol{\epsilon}_i.$$

Bei der Permutation der Daten muss die zeilenweise Zuordnung zum ersten Teil der Designmatrix bestehen bleiben, um die Einhaltung des Erwartungswertes unter der Nullhypothese zu gewährleisten. Daher muss der erste Teil der Designmatrix zu den Daten synchron mitpermutiert werden, so dass sich für den Vektor bei Permutation

$$\mathbf{y}_{\pi(i)} = \mathbf{x}_{1_{\pi(i)}} \mathbf{B}_1 + \mathbf{x}_{2_i} \mathbf{B}_2 + \boldsymbol{\epsilon}_{\pi(i)} \quad (2.39)$$

ergibt. Hier gilt somit nur $\mathbf{Y} | \mathbf{X}_1, \mathbf{X}_2 \stackrel{d}{=} \mathbf{Y}_\pi | \mathbf{X}_{1_\pi}, \mathbf{X}_2$ und daher wiederum

$$T(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2) \stackrel{d}{=} T(\mathbf{Y}_\pi, \mathbf{X}_{1_\pi}, \mathbf{X}_2). \quad (2.40)$$

Eine dazu inverse Permutation π^{-1} bewirkt entsprechend die Entstehung der ursprünglichen Reihenfolge der Zeilen der Beobachtungsmatrix und der Zeilen des ersten Teils der Designmatrix. Zugleich wird der für die Nullhypothese interessante Teil der Designmatrix invers permutiert, so dass

$$T(\mathbf{Y}_\pi, \mathbf{X}_{1_\pi}, \mathbf{X}_2) = T(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_{2_{\pi^{-1}}}). \quad (2.41)$$

Daher kann bei Verwendung der allgemeinen Nullhypothese (2.7) die Testproblematik auf ein geteiltes Modell zurückgeführt werden, wobei eine Permutation allein der hypothesenrelevanten Zeilen der Designmatrix \mathbf{X}_2 erfolgt, um die Permutationsverteilung

abzuleiten.

Nach dieser selektiven Permutation verfahren auch [Hummel *et al.* \(2008\)](#) bei der Untersuchung von Gengruppeneffekten. Auch hier bildet das geteilte lineare Modell (2.8) die Grundlage. Die Autoren stellen das gesamte Modell mit dem um den interessierenden Effekt reduzierten Modell in einer F -Statistik gegenüber und permutieren zur Bildung der Verteilungsfunktion dieser Teststatistik ebenfalls nur den interessierenden Teil der Designmatrix.

Eine weitere selektive Permutationsweise verfolgen [Kherad-Pajouh & Renaud \(2010\)](#). Um nur den interessierenden Effekt über eine Permutation zu testen, nutzen sie nicht die Permutation aller n Vektoren über die geteilte Designmatrix, sondern permutieren statt dessen die reduzierte Matrix $\tilde{\mathbf{Y}}$ vom Format $(n - s_1) \times p$, wie sie im folgenden Kapitel für Rotationsverfahren genutzt wird. Für die in dieser Arbeit durchgeführte Anpassung des Permutationsprinzips an korrelierte Stichprobenelemente mittels Blockpermutation (vgl. Kapitel 5) ist diese Variante aber nicht ideal, da die dort auftretende AR(1)-Struktur nur für die ursprünglichen $n \times p$ -dimensionalen Stichprobenvektoren gilt.

In der praktischen Umsetzung sind Permutationen selbst in Form der Monte-Carlo-Methode relativ aufwändig, vor allem wenn es sich dabei zudem um umfangreiche Simulationsrechnungen handelt wie im Kapitel 6 dieser Arbeit. Daher ist man in der programmiertechnischen Umsetzung bestrebt, sämtliche von der Permutation unbeeinflussten Größen außerhalb der Permutationsschritte zu bilden.

Bei Verwendung der stabilisierten multivariaten Verfahren (vgl. Abschnitt 2.3.2) als Teststatistiken innerhalb eines Permutationstests ist es beispielsweise möglich, die totale Produktsummenmatrix \mathbf{W} (vgl. Satz 2.7) außerhalb der Permutation zu berechnen. \mathbf{W} beinhaltet zwar die gesamte Designmatrix, da

$$\mathbf{W} = \mathbf{H} + \mathbf{G} = \mathbf{Y}'\mathbf{Q}_H\mathbf{Y} + \mathbf{Y}'\mathbf{Q}_G\mathbf{Y} = \mathbf{Y}'(\mathbf{Q}_H + \mathbf{Q}_G)\mathbf{Y}$$

mit $\mathbf{Q}_H = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}(\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C})^{-1}\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ und $\mathbf{Q}_G = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ aus (2.15) beziehungsweise (2.17). Man kann aber für das geteilte Modell zeigen, dass sich der zu permutierende Teil \mathbf{X}_2 der Designmatrix $\mathbf{X} = (\mathbf{X}_1\mathbf{X}_2)$ vollständig heraus kürzen lässt (siehe Anhang A).

2.3.4. Rotationstests

Anders als Permutationstests basieren Rotationstests auf einem parametrischen Prinzip, wobei sie die Eigenschaften der linkssphärischen Matrixverteilung ausnutzen. Doch auch hier wird ein P -Wert generiert, indem die Originalteststatistik ihrer Verteilung unter der Nullhypothese gegenübergestellt wird ([Langsrud, 2005](#)). Diese wiederum wird anhand der Teststatistiken für jeweils rotierte Daten ermittelt.

Die Anzahl möglicher Rotationen ist jedoch unbegrenzt und somit unabhängig vom Stichprobenumfang ([Kropf *et al.*, 2007](#)). Damit ist ein Rotationstest auch bei extrem geringen Stichprobenlängen anwendbar.

Grundlagen zur Rotation

Der Grundgedanke von Rotationstests ist, dass die Verteilung einer linkssphärisch verteilten Matrix $\tilde{\mathbf{Y}}$ (vgl. Abschnitt 2.1.2) gegenüber einer orthogonalen Rotation invariant ist. Ein Beispiel hierfür ist eine Matrix mit unabhängig, identisch normalverteilten Zeilenvektoren mit Erwartungswert $\mathbf{0}$, wie etwa in einem Einstichprobenproblem unter der Nullhypothese.

Durch wiederholt simulierte orthogonale Matrizen Δ können unbegrenzt viele Rotationen der linkssphärisch verteilten Matrix $\tilde{\mathbf{Y}}$

$$\tilde{\mathbf{Y}}_\delta = \Delta \tilde{\mathbf{Y}} \quad (2.42)$$

durchgeführt und damit nach Definition 2.5 die Rotationsverteilung einer Teststatistik bestimmt werden. Dabei wird eine entsprechende Teststatistik in jedem Rotationsschritt für die rotierte Matrix $\tilde{\mathbf{Y}}_\delta$ berechnet, wobei unter der Nullhypothese

$$T(\tilde{\mathbf{Y}}) \stackrel{d}{=} T(\tilde{\mathbf{Y}}_\delta) \quad (2.43)$$

gilt. Auch hier werden im Weiteren Teststatistiken betrachtet, die so bestimmt sind, dass sie für Abweichungen von der Nullhypothese größere Werte annehmen.

Da die Anzahl möglicher Rotationen im Gegensatz zu der bei Permutationen unbegrenzt ist, wird die Verteilung der Teststatistik unter der Nullhypothese unabhängig von der Anzahl der Stichprobenelemente anhand einer geeigneten Anzahl n_δ an Rotationen bestimmt. Analog zur Monte-Carlo-Simulation bei Permutationstests wird dieser Umfang in der Regel anhand der Fehlerabschätzung nach (2.33) zuvor festgelegt und sollte generell nicht unter $n_\delta = 1\,000$ liegen.

Entsprechend dem Prinzip bei Permutationstests kann die Nullhypothese schließlich abgelehnt werden, wenn die Teststatistik basierend auf den Originalwerten mindestens dem Wert des $(1 - \alpha)$ -Quantils der Rotationsverteilung entspricht.

Der P -Wert verkörpert wiederum das Verhältnis der Teststatistiken bei Rotation, die größer oder gleich der Originalteststatistik sind, zur Rotationsanzahl. Auch hier fließt die Originalstatistik in die Rotationsverteilung ein, so dass sich der P -Wert analog zu (2.32), dem der Permutation, aus

$$P = \frac{\#(T_\delta \geq T_0) + 1}{n_\delta + 1} \quad (2.44)$$

ergibt. Hierbei stellen T_0 die originale Teststatistik, T_δ die Teststatistiken an den rotierten Daten und n_δ die Anzahl der durchgeführten Rotationen dar.

Zumeist müssen die zu rotierenden Daten aber erst in die Form einer unter der Nullhypothese linkssphärisch verteilten Matrix gebracht werden (Kropf *et al.*, 2007).

Dazu wird wieder das lineare Modell (2.4) mit der allgemeinen Nullhypothese (2.7) in Form des geteilten linearen Modells (2.8)

$$\mathbf{Y} = (\mathbf{X}_1 \mathbf{X}_2) \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix} + \mathbf{E} = \mathbf{X}_1 \mathbf{B}_1 + \mathbf{X}_2 \mathbf{B}_2 + \mathbf{E}$$

mit der korrespondierenden Nullhypothese (2.9) $H_0 : \mathbf{B}_2 = \mathbf{0}$ betrachtet.

Die Nichtzentralität unter der Nullhypothese $E(\mathbf{Y}) = \mathbf{X}_1 \mathbf{B}_1$ wird aus den Daten durch Multiplikation mit der Projektionsmatrix $\mathbf{R}_1 = \mathbf{I}_n - \mathbf{X}_1(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1$ eliminiert. Der alleinigen Eliminierung folgt jedoch eine Rangreduzierung, da die Zeilen der Beobachtungsmatrix nicht mehr unkorreliert sind. Daher wird die Wurzel von \mathbf{R}_1 vom Format $n \times (n - s_1)$, genauer gesagt ihre Transponierte \mathbf{A}' , berechnet, wobei $\mathbf{A} \mathbf{A}' = \mathbf{R}_1$ und $\mathbf{A}' \mathbf{A} = \mathbf{I}_{n-s_1}$. Damit kann nunmehr durch

$$\tilde{\mathbf{Y}}_{(n-s_1) \times p} = \mathbf{A}'_{(n-s_1) \times n} \mathbf{Y}_{n \times p} \quad (2.45)$$

die linkssphärische zeilenreduzierte Beobachtungsmatrix $\tilde{\mathbf{Y}}$ gebildet werden. Dementsprechend ist $\tilde{\mathbf{Y}} \sim N_{(n-s_1) \times p}(\mathbf{0}, \mathbf{I}_{n-s_1} \otimes \Sigma)$ verteilt, da

$$\begin{aligned} E(\tilde{\mathbf{Y}}) &= E(\mathbf{A}'_{(n-s_1) \times n} \mathbf{Y}_{n \times p}) = \mathbf{A}' E(\mathbf{Y}) = \mathbf{I}_{n-s_1} \mathbf{A}' E(\mathbf{Y}) \\ &= \mathbf{A}' \mathbf{A} \mathbf{A}' E(\mathbf{Y}) = \mathbf{A}' \mathbf{R}_1 E(\mathbf{Y}) = \mathbf{0}, \end{aligned}$$

darauf basierend, dass

$$\begin{aligned} \mathbf{R}_1 E(\mathbf{Y}) &= (\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1) E(\mathbf{Y}) \\ &= (\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1) \mathbf{X}_1 \mathbf{B}_1 \\ &= \mathbf{X}_1 \mathbf{B}_1 - \mathbf{X}_1(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_1 \mathbf{B}_1 = \mathbf{0} \end{aligned}$$

und entsprechend Satz 2.2

$$\begin{aligned} \text{Var}(\tilde{\mathbf{Y}}) &= \text{Var}(\mathbf{A}'_{(n-s_1) \times n} \mathbf{Y}_{n \times p}) = (\mathbf{A}' \mathbf{I}_n (\mathbf{A}')') \otimes \Sigma \\ &= (\mathbf{A}' \mathbf{A}) \otimes \Sigma = \mathbf{I}_{n-s_1} \otimes \Sigma \end{aligned}$$

gilt.

Mit dieser Matrix $\tilde{\mathbf{Y}}$ lassen sich Rotationen nach (2.42) durchführen und die jeweilige Teststatistik der rotierten Daten $T(\tilde{\mathbf{Y}}_\delta)$ bilden. Diese ergeben dann nach (2.44) im Vergleich zu der Teststatistik der originalen linkssphärisch verteilten Daten $T(\tilde{\mathbf{Y}})$ den P -Wert.

Die im vorherigen Abschnitt erwähnte Arbeit von [Kherad-Pajouh & Renaud \(2010\)](#) nutzt ebenso diese Sphärizitätseigenschaft zur Begründung der Funktionalität ihres Permutationstests. Diese Permutation entspricht der Rotation (2.42), wenn Δ eine Permutationsmatrix ist, also eine Matrix mit nur genau einem 1-Eintrag in jeder Zeile und Spalte.

Eine zufällige Rotationsmatrix Δ vom Format $(n - s_1) \times (n - s_1)$ kann über die Orthogonalisierung

$$\Delta = \Delta^* (\Delta^* \Delta^*)^{-\frac{1}{2}} \quad (2.46)$$

einer $(n - s_1) \times (n - s_1)$ -dimensionalen Matrix mit unabhängigen standardnormalverteilten Einträgen Δ^* gebildet werden. In jedem Fall gilt dabei $\Delta' \Delta = \Delta \Delta' = \mathbf{I}_{n-s_1}$.

In manchen Fällen kann die Teststatistik $T(\mathbf{Y})$ so umgeformt werden, dass sie als Funktion $\tilde{T}(\tilde{\mathbf{Y}})$ der zeilenreduzierten Beobachtungsmatrix $\tilde{\mathbf{Y}}$ dargestellt wird, wobei $T(\mathbf{Y}) = \tilde{T}(\tilde{\mathbf{Y}})$ gilt. Dies ist beispielsweise beim modifizierten PC-Test (Ding *et al.*, 2012) der Fall. Damit kann die Generierung der Nullverteilung von $T(\mathbf{Y})$ anhand einer Folge zufälliger Rotationen von $\tilde{\mathbf{Y}}$ durchgeführt werden.

In anderen Fällen, wie den Tests auf Basis paarweiser Abstandsmaße der Stichproben (Kropf & Adolf, 2009, vgl. auch Abschnitt 2.3.2 dieser Arbeit) ist diese Umstellung der Teststatistik nur schwer möglich oder beeinträchtigt die Anschaulichkeit der Teststatistik. Hier kann man aus der rotierten reduzierten Matrix $\tilde{\mathbf{Y}}_\delta$ durch Multiplikation mit der Matrix \mathbf{A} wieder eine $n \times p$ -dimensionale Matrix \mathbf{Y}_δ erstellen

$$\mathbf{Y}_{\delta_{n \times p}} = \mathbf{A}_{n \times (n-s_1)} \tilde{\mathbf{Y}}_{\delta_{(n-s_1) \times p}}. \quad (2.47)$$

Die Multiplikation der reduzierten Matrix (2.47) mit dem Faktor \mathbf{A} hebt die ursprüngliche Reduzierung in gewisser Weise auf, da $\tilde{\mathbf{Y}} = \mathbf{A}\mathbf{A}'\mathbf{Y} = \mathbf{R}_1\mathbf{Y}$ (vgl. Abschnitt 2.2). Damit ist also lediglich die durch \mathbf{X}_1 verursachte Nichtzentralität eliminiert. Die verwendete Teststatistik muss deshalb invariant bezüglich dieser Nichtzentralität sein, beziehungsweise diese Projektion ohnehin als ersten Schritt beinhalten (wie in Kropf & Adolf, 2009). Die Folge der rotierten reduzierten Datenmatrizen $\tilde{\mathbf{Y}}_\delta$ wird dabei immer auf der Basis von $\tilde{\mathbf{Y}}$ erzeugt. Vor der Berechnung der Teststatistik erfolgt dann jeweils die Transformation der rotierten Daten nach (2.47).

In diesen beiden betrachteten Anwendungen des Rotationsprinzips können so jene Teststatistiken genutzt werden, die einem klassischen parametrischen Ansatz schwer zugänglich sind. Das Rotationsprinzip ermöglicht dabei die Berücksichtigung von Störfaktoren (in der Designteilmatrix \mathbf{X}_1 zusammengefasst) und die Anwendung bei kleinen Stichproben, bei denen Permutationstests aufgrund der zu geringen Anzahl möglicher Permutationen nicht mehr einsetzbar sind.

3. Modellbesonderheiten bei funktionellen MRT-Daten

Einige besondere Eigenschaften des linearen Modells bei Anwendung auf fMRT-Daten erfordern eine Anpassung der Datenauswertung. In diesem Kapitel wird nach einer Einführung in die funktionelle Magnetresonanztomographie zunächst auf diese Besonderheiten des Modells eingegangen. Anschließend wird anhand einfacher Simulationen demonstriert, wie sich die Korrelation der Stichprobenelemente auf den Fehler 1. Art auswirkt.

3.1. Grundlagen der funktionellen Magnetresonanztomographie

Die Magnetresonanztomographie (MRT) ist ein nicht-invasives Verfahren, dessen Grundlagen 1973 von [Lauterbur](#) sowie von [Mansfield & Grannell](#) gelegt wurden. Die MRT nutzt die magnetischen Eigenschaften von Protonen zur Erzeugung eines Signals, in erster Linie deren Spin, eine quantenmechanische Eigenschaft im Sinne eines Eigendrehimpulses. Im vom Tomographen erzeugten statischen äußeren Magnetfeld richten sich die Protonenspins parallel oder antiparallel zum Magnetfeld aus ([Siemens, 2003](#)). Größtenteils heben sich die parallelen Protonenspins mit den antiparallelen auf. Lediglich ein kleiner überschüssiger Teil paralleler Spins addiert sich zur sogenannten makroskopischen Magnetisierung. Dabei präzedieren die Spins um die Feldrichtung, das heißt, sie kreiseln mit einer von der Stärke des äußeren Magnetfeldes und der Art des Atoms abhängigen Präzessionsfrequenz, der sogenannten Larmorfrequenz, um ihre Drehachse. Ein summarisches und somit messbares MR-Signal wird aber erst erzeugt, indem die Spins durch einen Hochfrequenzpuls angeregt werden. Das daraus resultierende Signal ist protonenspezifisch, da der Hochfrequenzpuls nur auf die Spins mit entsprechender Larmorfrequenz wirkt und diese wiederum vom Atomkern abhängt.

Um aber ein Bild zu erstellen, ist darüber hinaus eine räumliche Zuordnung des MR-Signals notwendig. Dies geschieht durch Gradienten, die teilweise zeitgleich zum Hochfrequenzpuls in verschiedene Raumrichtungen geschaltet werden. Dabei handelt es sich um lokale Veränderungen der Feldstärke, die das statische äußere Magnetfeld innerhalb einer Schicht variieren und somit eine unterschiedliche Larmorfrequenz bei den angeregten Spins in dieser Schicht erzeugen. Anschließend wird durch eine Frequenzanalyse die zugehörige Signalstärke für jedes Voxel, die einzeln referenzierbaren Volumenelemente des Gehirns, ermittelt. Das MR-Bild wird konstruiert, indem diese Signale als Grauwert je Voxel dargestellt werden ([Siemens, 2003](#)).

Durch die hohe räumliche Auflösung, die anatomisch detaillierte Zuordnung der Informationen sowie die mögliche Wiederholbarkeit der Messungen in kurzen Abständen innerhalb einer Session, wie auch zu verschiedenen Sessions (Lazar, 2008) und nicht zuletzt wegen der Nicht-Invasivität des Verfahrens, eignet sich die Magnetresonanztomographie sehr gut für sogenannte funktionelle Aufnahmen.

Das meist verwendete Prinzip der funktionellen Bildgebung unter Nutzung des BOLD- (blood-oxygen-level-dependent) Effekts basiert darauf, dass der zerebrale Blutfluss einem körpereigenen Kontrastmittel für die MRT gleicht (Ogawa *et al.*, 1990). Bei einer Aktivierung von Neuronen verstärkt sich der zerebrale Blutfluss in der entsprechenden Region, um den Bedarf an Glukose und Sauerstoff abzudecken (Lazar, 2008). Damit erreicht ein Überschuss an Oxyhämoglobin – der an den Blutfarbstoff Hämoglobin gebundene Sauerstoff – die aktivierten Neurone. Hämoglobin wiederum enthält ein Eisenatom, das in Verbindung mit Sauerstoff die magnetischen Eigenschaften des Blutes beeinflusst, wodurch das MR-Signal bei erhöhtem Oxyhämoglobingehalt ansteigt. Anschließend ändert sich das Verhältnis von Oxy- zu Desoxyhämoglobin (Blutfarbstoff ohne Bindung an Sauerstoff) zugunsten des paramagnetischen Desoxyhämoglobins, was zum Abfall des MR-Signals führt (Schneider & Fink, 2007).

Diese Änderungen der magnetischen Eigenschaft des Blutes, aufgezeigt über die Veränderung des BOLD-Signals, werden anhand der sogenannten Hämodynamische Antwortfunktion (HRF, hemodynamic response function; vgl. Abschnitt 3.2.2) modelliert. Die Überkompensation an Sauerstoff zur neuronalen Aktivität ist zeitlich um einige Sekunden verzögert (Schneider & Fink, 2007) und auch der Abfall des MR-Signals vollzieht sich erst allmählich. Die Signaländerung macht zudem nur wenige Prozent des Gesamtsignals aus. Sie ist daher oftmals zu gering, um direkt anhand ihres Verlaufs erkennbar zu sein und wird im Allgemeinen erst durch statistische Analysen nachweisbar.

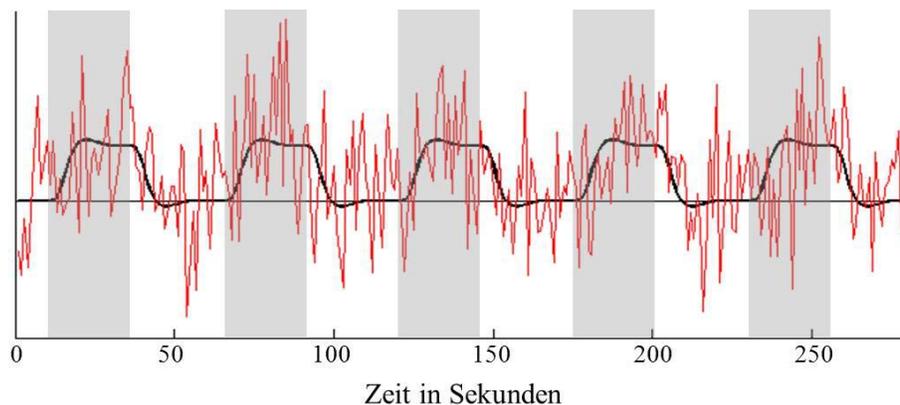


Abbildung 3.1.: Schematische Darstellung des Zeitverlaufs eines Experiments am Beispiel eines Voxels: wiederholt gegebener Reiz (grau), erwartetes Signal (schwarz) und gemessenes Signal (rot)

Abbildung 3.1 zeigt den Zusammenhang von Reiz, erwartetem und gemessenem Signal in einer schematischen Darstellung. Während der grau gekennzeichneten Zeitabschnitte wird ein Reiz gegeben. Das erwartete Signal (schwarz) ist zeitlich erkennbar vom

Reiz verschoben, steigt langsam an und fällt zwischen den Aktivierungen erst allmählich wieder ab. Das währenddessen tatsächlich gemessene Signal (rot) ist einem Rauschen unterlegen, so dass der Signalverlauf mitunter zwar erkennbar ist, vom erwarteten Signal aber teilweise stark abweichen kann.

Standardmäßig erfolgt die Auswertung des gemessenen Signals für jedes Voxel anhand eines univariaten allgemeinen linearen Modells unter Berücksichtigung der zeitlichen Korrelation (vgl. Kapitel 4).

Zur Auswertung von fMRT-Daten stehen etablierte Softwarepakete wie SPM, Statistical Parametric Mapping (SPM, 2009), als Toolbox von Matlab (Matlab, 2011) oder eigenständige Applikationen wie BrainVoyager (BrainInnovation, 2011) zur Verfügung.

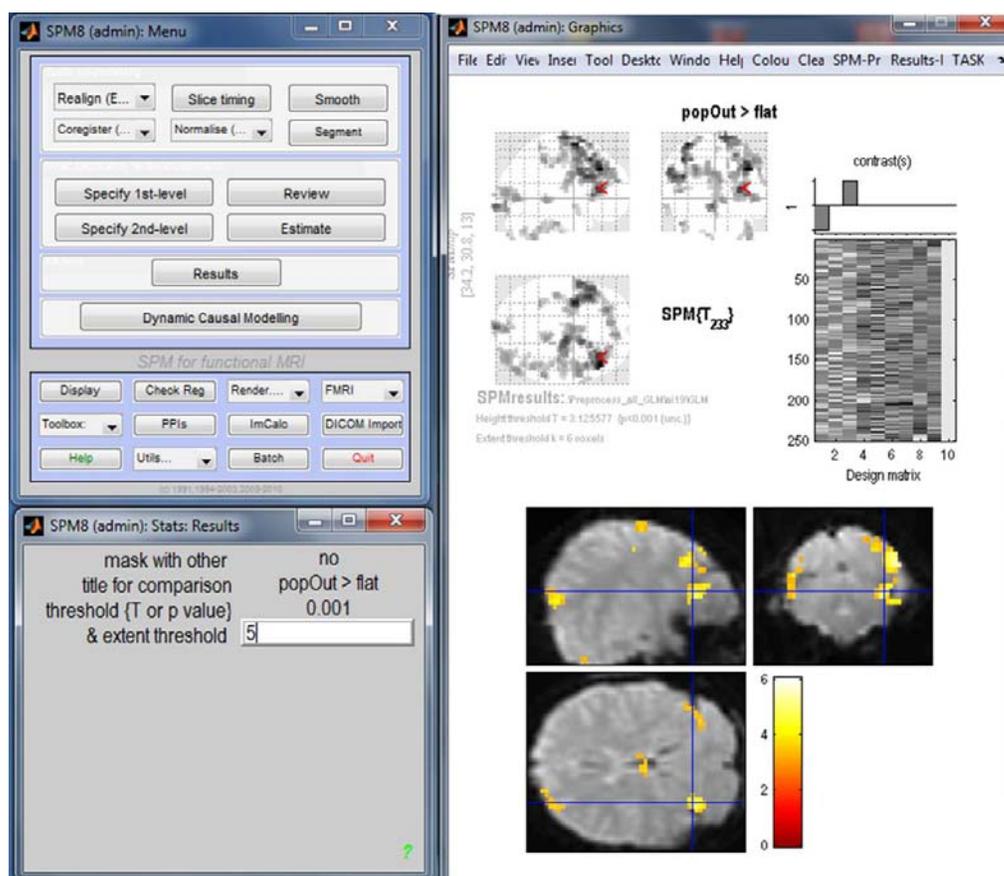


Abbildung 3.2.: Benutzeroberfläche der Toolbox SPM zur Auswertung von fMRT-Daten, Menü und Eingabemasken links, Anzeige der Designmatrix und des ausgewählten Kontrasts sowie Darstellung von Aktivierungen rechts (im Beispiel: first-level-Analyse, Differenz zweier Kontraste bei unadjustiertem Testniveau von 0,1 Prozent)

Abbildung 3.2 zeigt eine Aufnahme der Benutzeroberfläche der Toolbox SPM bei der Auswertung funktioneller MRT-Daten einer visuellen Studie (Baecke *et al.*, 2009).

Das Programm enthält verschiedene Funktionen zur Darstellung der voxelweisen Ergebnisse, wie die wahlweise Einstellung einer unadjustierten Irrtumswahrscheinlichkeit von $\alpha = 0,001$ oder die Verwendung von FWE- beziehungsweise FDR-Korrekturen für multiples Testen. Weiterhin ist die Erstellung der Designmatrix (vgl. Abschnitt 3.2.2) anhand des Programms intuitiv möglich. Zudem können verschiedene Kontraste einfach visualisiert werden.

Neben diesen Anwendungen bieten die Programme eine Unterstützung bei der Durchführung verschiedener Vorverarbeitungsschritte für funktionelle MRT-Daten an. Zusätzlich zur Bewegungskorrektur, die bereits während der Messung erfolgen und in die statistische Analyse einbezogen werden kann, sind dies eine Koregistrierung, die sogenannte slice-time-Korrektur, eine Normalisierung sowie eine räumliche Glättung.

Die Bewegungskorrektur der Daten ist notwendig, da es trotz sorgfältiger Positionierung und Fixierung nicht möglich ist, jegliche Kopfbewegung des Probanden zu unterdrücken (Schneider & Fink, 2007). Die Korrektur erfolgt standardmäßig auf die erste Aufnahme in Form von Translation und Rotation in jeweils alle drei Raumrichtungen. Sie wird an den Daten zumeist in Referenz zum ersten aufgenommenen Bild ausgeführt und ist wichtig, um die einzelnen Voxel im Zeitverlauf genau zu referenzieren. Bereits Lageänderungen im Millimeterbereich würden sonst die voxelweise Zuordnung der Messungen verfälschen. Die Informationen zur Bewegungskorrektur können bei der Datenanalyse zusätzlich in die Designmatrix einfließen, um mit der Bewegung assoziierte Aktivierungen in das Modell zu integrieren.

Zusätzlich kann eine Koregistrierung durchgeführt werden, welche die funktionellen Bilder auf höher aufgelöste anatomische Aufnahmen abgleicht (Schneider & Fink, 2007).

Die slice-time-Korrektur ist eine zeitliche Interpolation von Signalstärken und wird vor allem bei ereignisbezogenen Paradigmen benutzt (Frackowiak *et al.*, 2003). Durch die aufeinanderfolgende Aufnahme der einzelnen Schichten bei der typischen BOLD-fMRT-Bildgebung entsteht eine zeitliche Verzögerung zwischen der Aufnahme verschiedener Schichten – insbesondere, wenn viele Schichten gemessen werden. Die slice-time-Korrektur bewirkt indes eine Interpolation der Signale auf einen definierten Zeitpunkt zwischen zwei Messungen.

Die Normalisierung der Probandendaten auf ein standardisiertes Gehirn ist vor allem dann wichtig, wenn die Messergebnisse verschiedener Probanden im Rahmen einer second-level-Analyse gemeinsam analysiert werden. Denn die einzelnen Probanden unterscheiden sich in individueller Form und Größe des Gehirns (Schneider & Fink, 2007). Die Normalisierung bildet identische anatomische Strukturen aufeinander ab, um einen Vergleich von Aktivierungsmustern zu ermöglichen. Dabei wird ein von Talairach & Tournoux (1988) etabliertes Koordinatensystem verwendet, das beispielsweise in der Software SPM in einer vom MNI (Montreal Neurological Institute) weiterentwickelten Version implementiert ist (SPM, 2009).

Schließlich wird vor allem bei der univariaten Auswertung von fMRT-Daten häufig eine

räumliche Glättung des Signals anhand eines dreidimensionalen Gauß-Kerns durchgeführt. Dabei werden die Signale der einzelnen Voxel durch eine gewichtete additive Verknüpfung mit denen benachbarter Voxel geglättet (Friston *et al.*, 2007). Dies geschieht mit dem Ziel, zufällige Effekte herauszumitteln und somit einerseits die Sensitivität der voxelweisen Analyse zu erhöhen und andererseits die Vergleichbarkeit verschiedener Probanden zu verbessern (Schneider & Fink, 2007).

Bis auf die Bewegungskorrektur, die bereits während der Messungen oder als Vorverarbeitungsschritt realisiert wird und im Modell beachtet werden sollte, sind jegliche Vorverarbeitungsschritte für die first-level-Analyse nicht zwingend erforderlich, können aber der statistischen Datenauswertung vorgeschaltet werden. Daher wird im Weiteren nur auf die Besonderheiten der funktionellen MRT-Daten in Bezug auf das allgemeine lineare Modell eingegangen.

3.2. Besonderheiten des allgemeinen linearen Modells für fMRT-Daten

Die first-level-Analyse von fMRT-Daten wird für gewöhnlich anhand des allgemeinen linearen Modells $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$ analog zur Definition 2.6 durchgeführt (Frackowiak *et al.*, 2003). Dabei verkörpert die Beobachtungsmatrix \mathbf{Y} das BOLD-Signal, aufgenommen durch n Messungen (Stichprobenelemente) über jeweils p Voxel (Variablen).

Allerdings entspricht das Modell für fMRT-Daten nicht in allen Punkten den klassischen Annahmen. Im Folgenden werden die Besonderheiten des Modells für funktionelle Bilddaten beschrieben.

3.2.1. Format der Beobachtungsmatrix

In der Regel besteht eine funktionelle MRT-Session aus wiederholten Messungen unter verschiedenen Bedingungen, auf die in den Ausführungen zur Designmatrix genauer eingegangen wird.

Dabei kann das Signal für eine Vielzahl von Variablen (Voxel) gemessen werden. Die räumliche Auflösung innerhalb einer einzelnen Voxelschicht liegt in Abhängigkeit von der verwendeten Feldstärke zwischen 32×32 und 256×256 . Bei funktionellen Aufnahmen des gesamten Gehirns werden oftmals 32 Schichten gemessen. Zwar ist es möglich, noch mehr Schichten zu referenzieren, dadurch erhöht sich jedoch die Zeit zwischen zwei aufeinanderfolgenden Aufnahmen (Scans) ein- und derselben Voxelschicht. Bei 32 Schichten und einer entsprechenden Schichtdicke von etwa drei bis vier Millimetern liegen bei den derzeitigen Scannern je nach Experiment ein bis mehrere Sekunden zwischen zwei Scans (Friston *et al.*, 2007; SPM, 2009). Damit ist die Anzahl n der Aufnahmen, also der Stichprobenvektoren, bereits allein durch den zeitlichen Rahmen einer Session auf mehrere hundert Messungen begrenzt. Für das gesamte Gehirn referenziert man hingegen bei funktionellen Messungen beispielsweise oft $p = 131\,072$ ($64 \times 64 \times 32$) Voxel. Selbst wenn nur Aussagen über ein bestimmtes Hirnareal, eine sogenannte Region of

Interest (ROI), gemacht werden sollen, liegt die Anzahl der Variablen häufig weit über der der Stichprobenvektoren.

Aufgrund dieser hohen Dimension funktioneller Bilddaten, speziell da $n \ll p$, können klassische multivariate Testverfahren bereits allein wegen der auftretenden Singularität der Kovarianzmatrix nicht auf diese Daten angewendet werden.

3.2.2. Aufbau der Designmatrix

Die Designmatrix besteht bei Untersuchungen funktioneller MRT-Daten aus verschiedenen Einflussgrößen, von denen oft nur einige von eigenständigem Interesse sind. Die nicht direkt interessierenden Einflüsse sind beispielsweise Bewegungskorrekturen, die aufgrund geringer Lageänderungen des Probanden eingeschlossen werden, da diese mit Aktivierungen einhergehen können.

Ein Großteil der Einflussgrößen spiegelt vor allem den Versuchsaufbau der funktionellen Bildgebung wider. Werden dem Probanden verschiedene Reize präsentiert (egal, ob akustischer oder visueller Art oder als direkte Aufforderung, bestimmte Aufgaben zu absolvieren), sind diese je Reiz anhand eines dichotomen Stimulusvektors – des sogenannten Onsetvektors – darstellbar. Diese Vektoren enthalten ein binär kodiertes Paradigma, zu welchem Zeitpunkt der jeweilige Stimulus gegeben wird. Man unterscheidet dabei zwei Grundkonzepte: das Blockdesign, bei dem die Aktivierungsphase eine längere Zeitspanne andauert, und das ereignisbezogene Design, bei dem nur sehr kurze Reize gegeben werden. Die Aktivierungen sind jeweils durch Ruhephasen voneinander getrennt. Die Länge dieser Ruhebedingungen wird vor allem beim ereignisbezogenen Design zufällig um einige Sekunden variiert, damit sich der Proband zeitlich nicht gezielt auf die anschließend folgende Aktivierung einstellen kann. Beide Designs sind für jeweils einen Stimulusvektor in der Abbildung 3.3 schematisch dargestellt.

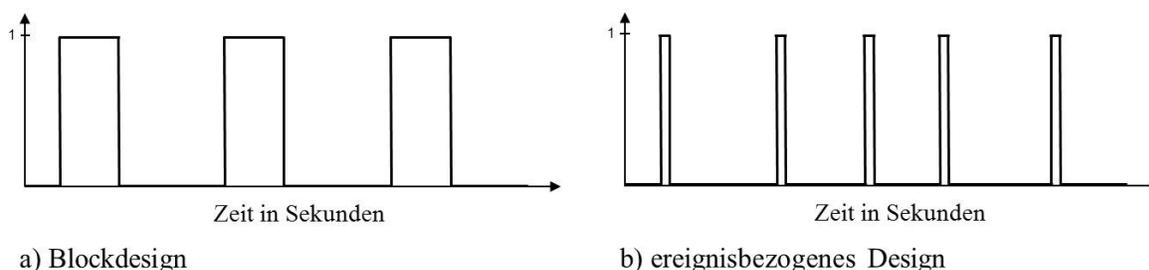


Abbildung 3.3.: Schematische Darstellung eines dichotomen Stimulusvektors für ein a) Blockdesign und ein b) ereignisbezogenes Design

Der Aufbau der Designmatrix für eine Untersuchung in der funktionellen Bildgebung wird durch Softwarepakete wie SPM oder BrainVoyager unterstützt (Friston *et al.*, 2007; BrainInnovation, 2011). Dabei kommt eine sogenannte Dummykodierung mit einer Konstanten zum Einsatz. Das bedeutet, dass in der Regel für jede Ausprägung einer zu

untersuchenden Einflussgröße ein separater Stimulusvektor gebildet wird. Für eine Einflussgröße mit zwei Ausprägungen entstünde beispielsweise die Kodierung

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & \vdots & \vdots \\ \vdots & 1 & 0 \\ \vdots & 0 & 1 \\ 1 & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \quad \mathbf{c} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix},$$

wobei die erste Spalte für das Gesamtmittel und die beiden weiteren Spalten für je eine Ausprägung stehen. Damit kann – wie in diesem Beispiel – die Voraussetzung von linear unabhängigen Spalten verletzt sein. Man könnte nun statt der Inversen $(\mathbf{X}'\mathbf{X})^{-1}$ eine verallgemeinerte Inverse verwenden oder die Paradigmen für die Designmatrix zunächst in eine Effektkodierung umformen

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & \vdots \\ \vdots & 1 \\ \vdots & -1 \\ 1 & \vdots \\ 1 & -1 \end{pmatrix} \quad \mathbf{c} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

um wieder linear unabhängige Spalten zu erhalten.

Mit dem jeweils angegebenen Kontrastgewichtsvektor \mathbf{c} lässt sich in beiden Fällen die Differenz der Ausprägungen testen (Fahrmeir *et al.*, 1996).

Wie im Kapitel 1 beschrieben, erfolgt die Veränderung des BOLD-Signals (als indirekten Nachweis der neuronalen Aktivität) mit einer gewissen zeitlichen Verzögerung. Die Paradigmen in ihrer ursprünglichen binären Form als Regressoren in die Designmatrix einfließen zu lassen, entspräche somit keiner realistischen Erwartung an die gemessenen Signale. Stattdessen muss die verzögerte, langsame Änderung des BOLD-Signals auf einen Reiz berücksichtigt werden. Dies geschieht, indem der jeweilige Onsetvektor \mathbf{o} mit der Hämodynamischen Antwortfunktion (HRF) gefaltet wird, welche die Veränderung des BOLD-Signals über die Zeit t bei Aktivierung modelliert.

Im Folgenden wird der zeitlich diskretisierte Stimulusvektor \mathbf{o} als Funktion der Zeit $o(t)$ dargestellt. Das Ergebnis der Faltung $x(t)$ fließt dann – wiederum als diskretisierter Vektor \mathbf{x} – als Regressor in die Designmatrix \mathbf{X} ein

$$x(t) = o * \text{HRF}(t) = \int_0^t o(t - \tau) \text{HRF}(\tau) d\tau, \quad (3.1)$$

wobei $*$ den Faltungsoperator darstellt (Friston *et al.*, 2007).

Dabei gibt es verschiedene Ansätze zur Approximation der HRF. Beim ersten Ansatz handelt es sich um eine einfache Gamma-Variate-Funktion nach [Boynton *et al.* \(1996\)](#). Damit kann der langsame Anstieg des BOLD-Signals sowie dessen anschließender Abfall bereits gut modelliert werden. Wenn kein weiterer Reiz gegeben wird, erfolgt der Abfall des Signals jedoch bis unter das Level der Signalstärke ohne Aktivierung. Dieser sogenannte post-stimulus undershoot lässt sich aber mit einer einzelnen Gamma-Variate-Funktion nicht modellieren, daher hat sich die Darstellung der HRF anhand der Differenz zweier Gamma-Variate-Funktionen bewährt und ist Standardeinstellung in den fMRT-Softwarepaketen ([SPM, 2009](#); [BrainInnovation, 2011](#)).

Die Approximation der Hämodynamischen Antwortfunktion erfolgt standardmäßig anhand der Doppel-Gamma-Variate-Funktion

$$\Gamma_{\text{HRF}}(t) = K_1 \left(\frac{t}{A_1} \right)^{\alpha_1} e^{-\frac{t-A_1}{\beta_1}} - K_2 \left(\frac{t}{A_2} \right)^{\alpha_2} e^{-\frac{t-A_2}{\beta_2}} \quad (3.2)$$

mit

$$\alpha_i = 8\ln(2) \frac{A_i^2}{W_i^2} \quad \text{und} \quad \beta_i = \frac{W_i^2}{8\ln(2)A_i}, \quad i = \{1, 2\}.$$

Dabei stellt A_i die Zeit bis zum Maximum der i -ten Gamma-Variate-Funktion in Sekunden dar, W_i ist die sogenannte full width at half maximum der i -ten Gamma-Variate-Funktion in Sekunden und K_i der entsprechende Formparameter.

Die Parametervoreinstellung im Softwarepaket SPM ([SPM, 2009](#)), die auch in den späteren Simulationen sowie dem Originalbeispiel verwendet wird, ist $A_1 = 6,00$, $W_1 = 5,10$, $K_1 = 1,00$, $A_2 = 16,00$, $W_2 = 8,00$ und $K_2 = 0,09$. Damit ergibt sich die in der Abbildung [3.4](#) dargestellte Modellierung der Γ_{HRF} , die der Veränderung des BOLD-Signals nach einer einmaligen Stimulation entspricht.

Die Faltung der originalen Onsetvektoren mit der HRF bewirkt, dass die Spalten der Designmatrix nicht mehr binär sind, sondern stetige Einträge aufweisen. Die Grenzen zwischen zwei Bedingungen sind nun fließend. Dadurch ist ein einfacher Zweistichprobenvergleich (Reiz- gegen Ruhebedingung), beispielsweise in Form eines Standard- t -Tests für zwei unabhängige Stichproben, nicht mehr ohne Weiteres möglich, sondern muss durch ein Regressionsproblem mit stetigen Einflussgrößen ersetzt werden.

Dennoch ist die Struktur der fMRT-Designmatrix im Rahmen des klassischen Modells realisierbar.

3.2.3. Kovarianzstruktur der Beobachtungen

Daten der funktionellen MRT sind in zweierlei Hinsicht korreliert. Zum einen weisen die Voxel (also die Variablen) eine räumliche Korrelation auf. Das erklärt sich aus ihrem räumlichen wie auch funktionellen Zusammenhang, was nicht zwingend identisch ist. Generell erfolgen Aktivierungen in zusammenhängenden Gebieten des Gehirns wie

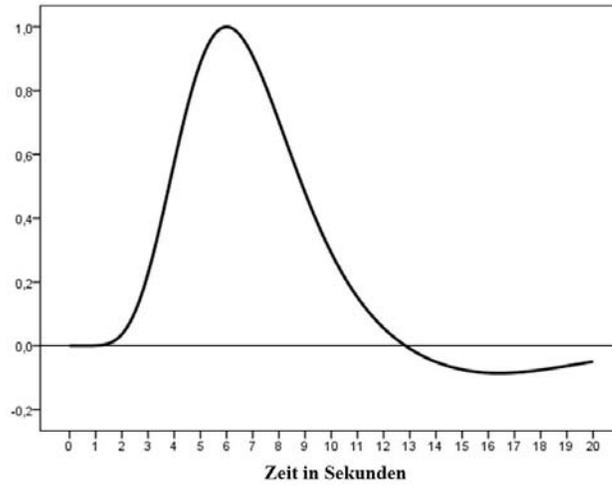


Abbildung 3.4.: Hämodynamische Antwortfunktion (Verlauf des BOLD-Signals) nach einmaliger Aktivierung zum Zeitpunkt 0, modelliert durch eine Doppel-Gamma-Variate-Funktion

zum Beispiel dem visuellen oder motorischen Kortex. Weiterhin werden viele Reize aber gleichzeitig in mehreren Gehirnregionen verarbeitet, was also funktionell korrelierten Gebieten entspricht. In jedem Fall können innerhalb der entsprechenden Regionen viele Voxel differenziert werden, die somit miteinander korrelierte Aktivierungen aufweisen. Diese räumliche Korrelation der Messungen entspricht den Korrelationen zwischen den verschiedenen Variablen des multivariaten linearen Modells und stellt somit kein grundsätzliches Problem für die Analyse dar. Zudem ist eine räumliche Glättung, wie sie oftmals für univariate Analysen erfolgt, nicht notwendig.

Zum anderen sind aber auch die einzelnen Messungen – die Stichprobenvektoren – voneinander abhängig, was sich zu einem großen Teil durch die verzögerte Veränderung des gemessenen BOLD-Signals bei einer Aktivierung erklären lässt. Diese zeitliche Korrelation verletzt die Voraussetzung von unabhängigen Stichprobenelementen. Das allgemeine lineare Modell nach Definition 2.6 muss dementsprechend in der Verteilungsannahme bezüglich der Kovarianzstruktur zu

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}, \quad \mathbf{E} \sim N_{n \times p}(\mathbf{0}, \mathbf{\Psi} \otimes \mathbf{\Sigma}) \quad (3.3)$$

erweitert werden, wobei $\mathbf{\Psi}$ die $n \times n$ -dimensionale Matrix der Kovarianzstruktur zwischen den Stichprobenvektoren verkörpert. Auf die genaue Struktur der zeitlichen wie auch räumlichen Kovarianzmatrizen wird in den folgenden Unterpunkten eingegangen.

Zeitliche Kovarianzstruktur der Beobachtungen

Für die zeitliche Kovarianzmatrix $\mathbf{\Psi}$ wird die Varianz für unterschiedliche Zeitpunkte (Stichprobenelemente) als konstant angenommen ($\psi_{ii}^2 = \psi^2 \forall i$) und somit bei der

Umrechnung in die Korrelationsmatrix

$$\mathbf{P} = [\text{Diag}(\Psi)]^{-\frac{1}{2}} \Psi [\text{Diag}(\Psi)]^{-\frac{1}{2}} \quad (3.4)$$

hier nur eine konstante Varianz ψ^2 herausgerechnet: $\Psi = \psi^2 \mathbf{P}$ (vgl. Definition 3.2). Die konstante Varianz ψ^2 wird nach (2.1) aus Gründen der Übersichtlichkeit und Eindeutigkeit in die Kovarianzmatrix Σ einbezogen, sodass sich für die Verteilung des Modells (3.3) die Kovarianzstruktur $\mathbf{P} \otimes \Sigma$ ergibt.

Diese zeitliche Korrelation der Residuen wird in der funktionellen Bildgebung vorwiegend anhand eines stationären autoregressiven Prozesses erster Ordnung (AR(1)-Prozess) modelliert (Frackowiak *et al.*, 2003), eine Annahme, die auch in dieser Arbeit getroffen wird. Der autoregressive Prozess ist dabei ein spezieller stochastischer Prozess entsprechend den folgenden Definitionen (Schlittgen & Streitberg, 2001).

Definition 3.1 *Ein stochastischer Prozess in diskreter Zeit ist eine Folge $(\mathbf{u}_t)_{t \in \mathbb{Z}}$ von Stichprobenvektoren \mathbf{u}_t . Dabei ist \mathbb{Z} die Menge der ganzen Zahlen.*

Definition 3.2 *Ein stochastischer Prozess $(\mathbf{u}_t)_{t \in \mathbb{Z}}$ heißt schwach stationär, wenn er mittelwert- und kovarianzstationär ist. Dabei bedeutet mittelwertstationär, dass für alle t gilt: $E(\mathbf{u}_t) = \mu_t = \mu$. Kovarianzstationarität drückt aus, dass die Kovarianzfunktion des Prozesses $\text{Cov}(\mathbf{u}_s, \mathbf{u}_t) = \gamma(s, t) = \gamma(s - t)$ für alle $s, t \in \mathbb{Z}$ allein von der Entfernung $s - t$ und nicht von deren Position innerhalb der Messreihe abhängt. Ein kovarianzstationärer Prozess ist zudem varianzstationär $\text{Var}(\mathbf{u}_t) = \sigma_t^2 = \sigma^2$.*

Im Falle normalverteilter Daten ergibt sich aus der schwachen automatisch die strenge Stationarität, welche die Invarianz der Verteilungsfunktion gegenüber Zeitverschiebungen definiert (Schlittgen & Streitberg, 2001), da die Verteilung solcher Daten vollständig über die ersten beiden Momente bestimmt ist. Aufgrund der in dieser Arbeit getroffenen Normalverteilungsannahme der Daten wird weiterhin lediglich der Begriff der Stationarität verwendet.

Bei einem stationären Prozess ist sowohl die Kovarianzfunktion $\gamma(s, t) = \gamma(s - t)$ als auch die zugehörige Korrelationsfunktion $\rho(s, t) = \rho(s - t)$ eine gerade Funktion bezüglich des sogenannten Lags $\tau = s - t$, des Abstands der Zeitpunkte, das heißt $\gamma(\tau) = \gamma(-\tau)$ und $\rho(\tau) = \rho(-\tau)$.

Weiterhin sind $\gamma(0) = \sigma^2$ und $\rho(0) = 1$. Die Beziehung beider Funktionen ist durch $\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$ definiert.

Definition 3.3 *Ein stochastischer Prozess (\mathbf{u}_t) heißt autoregressiver Prozess erster Ordnung, kurz AR(1)-Prozess, wenn er in der Form*

$$\mathbf{u}_t = \alpha \mathbf{u}_{t-1} + \mathbf{e}_t, \quad t \in \mathbb{Z} \quad (3.5)$$

dargestellt werden kann, wobei (\mathbf{e}_t) eine Folge unabhängig identisch verteilter Stichprobenvektoren ist.

Der AR(1)-Prozess ist speziell dann stationär, wenn $|\alpha| < 1$.

Wird also die zeitliche Korrelationsstruktur der Residuen der Messungen als Realisierung eines stationären AR(1)-Prozesses betrachtet, so hängt der Scan zum Zeitpunkt t unmittelbar vom Scan des vorhergehenden Zeitpunkts $t - 1$ ab. Die Korrelationsstruktur der Werte dieses Prozesses lässt sich mit Hilfe der Yule-Walker-Gleichung beschreiben, die eine iterative Berechnung der Autokorrelationsfunktion ermöglicht (Schlittgen & Streitberg, 2001).

Ohne Beschränkung der Allgemeinheit sei $E(\mathbf{u}_t) = 0$. Die Definitionsgleichung $\mathbf{u}_t = \alpha\mathbf{u}_{t-1} + \mathbf{e}_t$ wird mit $\mathbf{u}_{t-\tau}$, $\tau > 0$ multipliziert und der Erwartungswert des Produkts gebildet. Dabei ist τ der sogenannte Lag, der Abstand der Zeitpunkte. Es ergibt sich

$$\begin{aligned} E(\mathbf{u}_{t-\tau}\mathbf{u}_t) &= E(\mathbf{u}_{t-\tau}(\alpha\mathbf{u}_{t-1} + \mathbf{e}_t)) \\ &= \alpha E(\mathbf{u}_{t-\tau}\mathbf{u}_{t-1}) + E(\mathbf{u}_{t-\tau}\mathbf{e}_t) \end{aligned}$$

Wegen $E(\mathbf{u}_t) = 0$ und damit $E(\mathbf{u}_{t-\tau}\mathbf{u}_t) = \text{Cov}(\mathbf{u}_{t-\tau}, \mathbf{u}_t) = \gamma(\tau)$ und $E(\mathbf{u}_{t-\tau}\mathbf{e}_t) = 0$ folgt

$$\gamma(\tau) = \alpha\gamma(\tau - 1),$$

und da $\rho(\tau) = \frac{\gamma(\tau)}{\sigma^2}$, resultiert daraus schließlich die Yule-Walker-Gleichung des AR(1)-Prozesses

$$\rho(\tau) = \alpha\rho(\tau - 1).$$

Mit $\rho(0) = 1$ ergibt sich eine Autokorrelation von

$$\rho(\tau) = \alpha^{|\tau|}. \quad (3.6)$$

Satz 3.1 *Mit (3.6) lässt sich die Korrelationsmatrix \mathbf{P} für das Modell (3.3) schreiben als*

$$\mathbf{P} = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \ddots & \vdots \\ \rho^2 & \rho & 1 & \ddots & \rho^2 \\ \vdots & \ddots & \ddots & \ddots & \rho \\ \rho^{n-1} & \cdots & \rho^2 & \rho & 1 \end{pmatrix}. \quad (3.7)$$

Somit sind zeitlich nahe Messungen hoch miteinander korreliert, während die Korrelation zwischen weiter entfernten Scans abnimmt.

Dabei ist zudem vorteilhaft, dass die Schätzung der zeitlichen Korrelation anhand der Schätzung nur eines Parameters, des Korrelationskoeffizienten ρ , realisiert werden kann.

Räumliche Kovarianzstruktur der Beobachtungen

Bei der in den Softwarepaketen standardmäßig implementierten univariaten Auswertung des gemessenen BOLD-Signals ist die räumliche Kovarianzstruktur nur von sekundärer Bedeutung. Bei der multiplen Adjustierung der voxelweisen Ergebnisse nutzen einige Verfahren (z. B. Bonferroni) die Korrelationsstruktur der einzelnen Testvariablen nicht aus und erzielen so eine geringere Güte. Daher werden FWE- und FDR-Korrekturen benutzt, die dies berücksichtigen (vgl. Kapitel 1).

Bei multivariater Betrachtung wird die räumliche Kovarianzstruktur ausgenutzt, um beispielsweise Hauptkomponenten zu bilden. Eine vorherige genaue Kenntnis der Struktur ist dabei jedoch nicht nötig. Die multivariate Auswertung wird dabei zwar durch die Kovarianz der Spalten beeinflusst, eine Erweiterung des Modells analog zu der hier im Vordergrund stehenden Korrelation der Stichprobenvektoren ist aber nicht notwendig. Dennoch soll vor allem in den im Kapitel 6 folgenden Simulationen eine der fMRT-Analyse entlehnte räumliche Kovarianz berücksichtigt werden. Hierbei wird als Baustein eine sogenannte Compound-Symmetry-Struktur genutzt, die im Folgenden erklärt werden soll (Fahrmeir *et al.*, 1996; Davis, 2002).

Definition 3.4 *Bei einer Compound-Symmetry-Struktur besitzen alle Variablen gleiche Varianzen wie auch gleiche (positive) Kovarianzen.*

Man kann sich die Entstehung einer Compound-Symmetry-Struktur so vorstellen, dass jede Variable als Summe von zwei unabhängigen Komponenten entsteht, wobei die erste Komponente mit der Varianz σ_u^2 bei allen Variablen identisch, die zweite Komponente mit der Varianz σ_v^2 für jede Variable spezifisch ist.

Bei Betrachtung des Stichprobenvektors $\mathbf{y} = (y_1, \dots, y_p)$ gilt somit $\text{Var}(y_i) = \sigma_u^2 + \sigma_v^2$, $i = 1, \dots, p$, und $\text{Cov}(y_i, y_j) = \sigma_u^2$, $\forall i \neq j$.

Daraus ergibt sich eine Kovarianzstruktur \mathbf{CS} , die sich mit $\rho_{CS} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}$ als Korrelationsmatrix schreiben lässt:

$$\mathbf{CS} = (\sigma_u^2 + \sigma_v^2) \begin{pmatrix} 1 & \rho_{CS} & \cdots & \rho_{CS} \\ \rho_{CS} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{CS} \\ \rho_{CS} & \cdots & \rho_{CS} & 1 \end{pmatrix}. \quad (3.8)$$

Unter dieser räumlichen Kovarianzstruktur sind alle Variablen gleichstark miteinander korreliert, unabhängig davon, ob sie räumlich gesehen direkte Nachbarn sind oder weit voneinander entfernt liegen.

Dies wird für die Simulationen abgeschwächt, indem die Compound-Symmetry-Struktur innerhalb einzelner Variablenblöcke angenommen wird. Eine Block-Compound-Symmetry-Struktur entsteht dann mit

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{CS} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{CS} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{CS} \end{pmatrix}, \quad (3.9)$$

wobei die Variablen innerhalb der Blöcke gleich korreliert, die Blöcke aber unabhängig voneinander sind.

3.3. Notwendigkeit einer Anpassung an die zeitliche Korrelation

Die räumliche Korrelation der Daten verursacht keinerlei konzeptionelle Probleme bei der Analyse des allgemeinen linearen Modells – sie wird im multivariaten Fall sogar ausgenutzt. Die zeitliche Abhängigkeitsstruktur der Stichprobenvektoren hingegen verletzt die Voraussetzungen der klassischen Statistik und bewirkt ein Nicht-Einhalten des nominalen Testniveaus. Erste systematische Simulationsuntersuchungen dazu wurden in [Kose \(2007\)](#) dargestellt.

Um die Verletzung des Fehlers 1. Art, des Anteils der fälschlicherweise verworfenen Nullhypothesen, durch die Korrelation der Stichprobenelemente schon in diesem Abschnitt kurz zu verdeutlichen, wird ein einfacher Zweigruppenvergleich unter der Nullhypothese simuliert und der Fehler 1. Art bei Anwendung zweier der in [Abschnitt 2.3](#) vorgestellten Verfahren kontrolliert. Dies geschieht in Abhängigkeit von der Korrelation ρ innerhalb der Stichprobenelemente wie auch vom Stichprobenumfang n und der Variablenanzahl p mit einer nominalen Irrtumswahrscheinlichkeit von jeweils $\alpha = 0,05$. Begründet durch die Simulationsanzahl von je 10 000 Wiederholungen können Abweichungen innerhalb des Intervalls $[0,0457; 0,0543]$ als rein zufällig betrachtet werden.

Exemplarisch werden der Standardisierte Summentest sowie der skalenabhängige Principal Component Test zur Auswertung genutzt. Beide sind uneingeschränkt für alle Kombinationen von Stichprobenumfängen $n_i = \{25, 50, 100\}$, $i = 1, 2$, und Variablenanzahlen $p = \{1, 25, 500\}$ berechenbar. Dementsprechend sind Sonderfälle wie ein univariater Test enthalten.

Die Daten werden erzeugt, indem zunächst eine standardnormalverteilte $(n \times p)$ -dimensionale Matrix \mathbf{Y}_0 simuliert wird. Dieser wird dann eine zeitliche Korrelationsstruktur \mathbf{P} und eine räumliche Struktur $\mathbf{\Sigma}$ aufgeprägt. Die Korrelationsstruktur innerhalb der Stichprobenvektoren entspricht einer autoregressiven Struktur erster Ordnung der Form [\(3.7\)](#) mit $\rho \in \{0, 0; 0, 2; 0, 4; 0, 6; 0, 8\}$, die Abhängigkeitsstruktur der Variablen entspricht der Form [\(3.8\)](#), also einer einfachen Compound-Symmetry-Struktur mit $\rho_r = 0,5$. Diese Einarbeitung der Abhängigkeiten geschieht durch Multiplikation der symmetrischen Wurzeln von \mathbf{P} und $\mathbf{\Sigma}$ von links beziehungsweise rechts an die unkorrelierten Daten (vgl. [Satz 2.2](#))

$$\mathbf{Y} = \mathbf{P}^{\frac{1}{2}} \mathbf{Y}_0 \mathbf{\Sigma}^{\frac{1}{2}},$$

wobei die Wurzeln der Matrizen beispielsweise über ihre Eigenwertzerlegung gewonnen werden können (Details zur Simulation im [Abschnitt 6.1](#)).

Die Designmatrix besteht in diesem Beispiel lediglich aus einem Gesamtmittel und einer

Spalte mit Eins-Einträgen für die ersten n_1 Zeilen und Null-Einträgen für die anschließenden n_2 Zeilen. Getestet wird der Einfluss der zweiten Spalte der Designmatrix auf die simulierten Daten unter der Nullhypothese.

Für den Standardisierten Summentest (vgl. Abschnitt 2.3.2) wird für alle Parameterkonstellationen ein eindimensionaler Score bestimmt und für verschiedene Werte der zeitlichen Korrelation ρ die Einhaltung des Fehlers 1. Art beim Test mit diesem Score untersucht. Beim Principal Component Test hingegen wird – mit Ausnahme der univariaten Simulationen und wenn nicht a priori festgelegt – zunächst die Dimension q des Scores aus den Daten bestimmt. Dies geschieht in jeder Simulation separat anhand des inversen Jolliffe-Kriteriums (2.22). Anschließend erfolgt auch hier die Bestimmung des empirischen Fehlers 1. Art in Abhängigkeit des zeitlichen Korrelationskoeffizienten. Zusätzlich kommt auch ein PC-Test mit festgelegtem $q = 1$ für alle Parameterkonstellationen zum Einsatz. Innerhalb der stabilisierten Verfahren wird jeweils eine Wilk's Λ -Teststatistik genutzt.

Die Ergebnisse der Simulationen unter der Nullhypothese sind für beide Teststatistiken und die verschiedenen Parametereinstellungen in Tabelle 3.1 dargestellt.

Im eindimensionalen Fall ($p = 1$) entsprechen die Resultate beider Teststatistiken denen eines klassischen t -Tests und sind daher nur für den Standardisierten Summentest aufgeführt. Bei Unabhängigkeit kann das nominale Testniveau von $\alpha = 0,05$ hier – wie auch bei allen Dimensionen p der simulierten Daten – bei Durchführung des Standardisierten Summentests eingehalten werden. Mit wachsender Korrelation zwischen den Stichprobenvektoren erhöht sich der empirische Fehler 1. Art jedoch stark, so dass bei $\rho = 0,4$ schon jede fünfte, bei $\rho = 0,8$ etwa 50 Prozent der wahren Nullhypothesen durch den Standardisierten Summentest fälschlicherweise abgelehnt werden. Dabei variieren die Ergebnisse für diese Version der stabilisierten Testverfahren für die unterschiedlichen Stichprobenumfänge und Variablen kaum, da die Daten stets auf einen eindimensionalen Score zurückgeführt werden.

Bei Verwendung des Principal Component Tests soll einerseits der eindimensionale Score für alle Parameterkonstellationen, andererseits der über das inverse Jolliffe-Kriterium (2.22) bestimmte q -dimensionale Score betrachtet werden, um so – gesondert von der Dimension der Ausgangsdaten – auch den Einfluss der Scoredimension auf den Fehler 1. Art betrachten zu können. In jedem Fall kann das nominale Testniveau für alle Parameterkonstellationen bei unabhängigen Stichprobenvektoren ($\rho = 0,0$) eingehalten werden. Für korrelierte Daten steigt der empirische Fehler 1. Art auch bei diesem Testverfahren innerhalb jeder Simulationsreihe mit zunehmender Abhängigkeit an.

Beim eindimensionalen PC-Test erfolgt eine zunehmende Verletzung des Fehlers 1. Art mit größer werdender Korrelation der Stichprobenvektoren in ähnlicher Stärke wie beim Standardisierten Summentest. Auch hier wird für die höchste der simulierten Korrelationen ($\rho = 0,8$) jede zweite Nullhypothese fälschlicherweise abgelehnt, wobei die Ergebnisse für die verschiedenen Dimensionen der Ausgangsdaten ebenfalls kaum variieren. Bestimmt man die Dimension des PC-Tests aus den Daten, verändert sich diese in Ab-

Testverfahren	n	p	ρ				
			0,0	0,2	0,4	0,6	0,8
Standard.	50	1	0,053	0,104	0,195	0,326	0,503
Summentest	100	1	0,050	0,104	0,202	0,324	0,522
	200	1	0,048	0,107	0,197	0,322	0,516
	50	25	0,049	0,108	0,193	0,311	0,485
	100	25	0,050	0,111	0,191	0,323	0,511
	200	25	0,053	0,104	0,199	0,321	0,505
	50	500	0,050	0,108	0,196	0,317	0,493
	100	500	0,051	0,107	0,196	0,321	0,509
	200	500	0,051	0,107	0,197	0,327	0,505
Skalenabh.	50	25	0,054	0,120	0,217	0,354	0,571
PC-Test	100	25	0,052	0,116	0,203	0,345	0,548
eindimensional	200	25	0,054	0,108	0,207	0,330	0,522
	50	500	0,055	0,128	0,219	0,357	0,579
	100	500	0,053	0,111	0,209	0,343	0,549
	200	500	0,052	0,110	0,201	0,336	0,526
	Skalenabh.	50	25	(1) 0,050	(1) 0,153	(2) 0,420	(3) 0,801
PC-Test	100	25	(1) 0,049	(1) 0,113	(1) 0,206	(1) 0,499	(3) 0,933
q -dimensional	200	25	(1) 0,053	(1) 0,106	(1) 0,204	(1) 0,328	(2) 0,719
	50	500	(1) 0,049	(1) 0,114	(1) 0,466	(5) 0,996	(6) 1,00
	100	500	(1) 0,051	(1) 0,110	(3) 0,768	(11) 1,00	(11) 1,00
	200	500	(1) 0,051	(2) 0,311	(12) 0,992	(22) 1,00	(21) 1,00

Tabelle 3.1.: Empirischer Fehler 1. Art des Standardisierten Summentests sowie des Principal Component Tests bei korrelierten Stichprobenelementen in Abhängigkeit vom zeitlichen Korrelationskoeffizienten ρ , dem Stichprobenumfang n und der Variablenanzahl p ($\alpha = 0,05$, 10 000 Wiederholungen, (\tilde{q}) - Median der Dimension q des PC-Tests für die jeweiligen Simulationen)

hängigkeit der Originaldimensionen und der zeitlichen Korrelation. Bei größerer Abhängigkeit der Stichprobenvektoren vergrößert sich auch die Dimension der berechneten Scores. Verstärkt wird dies mit zunehmender Variablenanzahl. Bei $n = 200$, $p = 25$ und $\rho = 0,0$ wird im Median ein eindimensionaler Score gebildet, was bei der hier unterlegten Compound-Symmetry-Struktur der Variablen zu erwarten ist. Bei $\rho = 0,8$ werden bereits zweidimensionale Linearkombinationen erzeugt. Liegt die Variablenanzahl unterhalb der Stichprobengröße, verringert sich die Größe der geschätzten Scoredimension mit größer werdendem Stichprobenumfang. Liegt die Variablenanzahl jedoch über der Stichprobengröße, vergrößert sich die Dimension q der Scores mit zunehmender zeitli-

cher Korrelation wie auch mit größer werdendem Stichprobenumfang, was in Tabelle 3.1 anhand der medianen Dimension \tilde{q} dargestellt wird.

Mit Anstieg der Dimension des Principal Component Tests steigt die Verletzung des Fehlers 1. Art stark an. Bei großer Variablenanzahl und starker zeitlicher Korrelation – was selbst bei einer einfachen räumlichen Korrelationsstruktur zu zweistelligen Dimensionsgrößen führt – wird letztlich jede Nullhypothese fälschlicherweise verworfen.

Letztlich ergibt sich somit bei allen Stichprobenumfängen, Variablenanzahlen und Teststatistiken zwingend die Notwendigkeit der Beachtung einer Abhängigkeitsstruktur der Stichprobenelemente, um die Einhaltung des nominalen Testniveaus gewährleisten zu können.

4. Etablierte Anpassungsstrategien für korrelierte Stichprobenelemente

Da die Entwicklung von Analyseverfahren für den Umgang mit zeitlich korrelierten fMRT-Daten ursprünglich für das univariate Modell erfolgte, wird im Unterkapitel 4.1 zuerst auf die Anpassungsstrategien für das univariate lineare Modell eingegangen. Im Anschluss wird deren Entsprechung im multivariaten Modell betrachtet. Danach werden diese parametrischen Verfahren sowie weitere Lösungsansätze kurz diskutiert. Eine Alternative zur parametrischen Adaption der Testverfahren an die zeitliche Korrelation wird in Kapitel 5 betrachtet.

4.1. Univariate Anpassungsstrategien für zeitlich korrelierte Daten

Von Beginn der fMRT-Analysen an wurden Verfahren benutzt und weiterentwickelt, die die zeitliche Korrelation der Messungen berücksichtigen (Friston *et al.*, 2007). Darunter fallen unter anderem die Berücksichtigung der Korrelation innerhalb der Teststatistik, die in Abschnitt 4.1.1 vorgestellt wird, sowie ein Pre-Whitening (Abschnitt 4.1.2), das die Abhängigkeiten aus den Residuen eliminiert.

Analog zu (2.10) wird das univariate Modell bei abhängigen Stichprobenelementen betrachtet

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{P}), \quad (4.1)$$

wobei \mathbf{P} die $n \times n$ -dimensionale zeitliche Korrelationsmatrix in Form eines AR(1)-Prozesses darstellt (vgl. Satz 3.1).

Bei der Ableitung der Verfahren wird jeweils von einem bekannten Korrelationskoeffizienten ρ und dementsprechend einer bekannten zeitlichen Korrelationsstruktur \mathbf{P} ausgegangen. In den Anwendungen werden dann Schätzungen für den Parameter eingesetzt.

4.1.1. Satterthwaite-Approximation

Der erste Ansatz, die Autokorrelation der Stichprobenvektoren zu beachten, ist die Einbeziehung der Abhängigkeit in die t -Teststatistik (SPM, 2009). Dies wird über eine Varianzkorrektur und eine Satterthwaite-Approximation im Freiheitsgrad der t -Verteilung

berücksichtigt (Frackowiak *et al.*, 1997).

Die OLS-Schätzung $\hat{\beta}$ des Parametervektors (2.11) verändert sich bei Korrelation der Daten nicht. Die Kovarianzstruktur der Daten geht jedoch entsprechend in die Varianz des Parameterschätzers ein

$$\begin{aligned}\Sigma_{\hat{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{Y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2\mathbf{X}^-\mathbf{P}\mathbf{X}^{-'},\end{aligned}$$

wobei $\mathbf{X}^- = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ die Moore-Penrose-Inverse der Designmatrix \mathbf{X} verkörpert. Aufgrund der veränderten Varianz des Parameterschätzers ändert sich die t -Teststatistik (2.12) zu

$$t = \frac{\mathbf{c}'\hat{\beta}}{\sqrt{\hat{\sigma}^2\mathbf{c}'(\mathbf{X}^-\mathbf{P}\mathbf{X}^{-'})\mathbf{c}}} \sim t_\nu. \quad (4.2)$$

Darin kann die Varianz σ^2 durch

$$\hat{\sigma}^2 = \frac{\mathbf{y}'\mathbf{R}\mathbf{y}}{\text{Spur}(\mathbf{R}\mathbf{P})} \quad (4.3)$$

geschätzt werden, wobei $\mathbf{R} = \mathbf{I} - \mathbf{X}\mathbf{X}^-$ die residuenbildende Matrix ist (vgl. weiter unten).

Zähler und Nenner der modifizierten Teststatistik (4.2) sind auch jetzt noch unabhängig voneinander, der Nenner aber ist nun nicht mehr exakt χ^2 -verteilt. Aus diesem Grund wird der Freiheitsgrad anhand einer Satterthwaite-Approximation hergeleitet, indem die ersten beiden Momente der χ^2 -Verteilung mit denen der wahren Verteilung gleichgesetzt werden. Die Momente der Residualabweichungen $\text{E}(\boldsymbol{\epsilon}'\boldsymbol{\epsilon})$ und $\text{Var}(\boldsymbol{\epsilon}'\boldsymbol{\epsilon})$ lassen sich – wie bereits die Varianzschätzung (4.3) – mit Hilfe der quadratischen Form ausdrücken, denn mit $\hat{\beta}$ aus (2.11) gilt

$$\begin{aligned}\boldsymbol{\epsilon} &= \mathbf{y} - \mathbf{X}\hat{\beta} \\ &= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \\ &= \mathbf{R}\mathbf{y},\end{aligned}$$

und somit $\boldsymbol{\epsilon}'\boldsymbol{\epsilon} = \mathbf{y}'\mathbf{R}\mathbf{y}$, da \mathbf{R} idempotent ist. Für die quadratische Form gilt

$$\begin{aligned}\text{E}(\mathbf{y}'\mathbf{R}\mathbf{y}) &= \text{E}(\text{Spur}(\mathbf{y}'\mathbf{R}\mathbf{y})) \\ &= \text{E}(\text{Spur}(\mathbf{R}\mathbf{y}\mathbf{y}')) \\ &= \text{Spur}(\mathbf{R}\text{E}(\mathbf{y}\mathbf{y}')) \\ &= \text{Spur}(\mathbf{R}(\text{Var}(\mathbf{y}) + \text{E}(\mathbf{y})\text{E}(\mathbf{y}')))) = \text{Spur}(\mathbf{R}\sigma^2\mathbf{P}) \\ &= \sigma^2\text{Spur}(\mathbf{R}\mathbf{P}),\end{aligned}$$

was die Varianzschätzung (4.3) erklärt, und entsprechend

$$\text{Var}(\mathbf{y}'\mathbf{R}\mathbf{y}) = 2\sigma^4\text{Spur}((\mathbf{R}\mathbf{P})(\mathbf{R}\mathbf{P}))$$

(Beweis siehe [Seber, 1977](#)).

Die ersten beiden Momente einer χ_ν^2 -Verteilung sind $E(\chi_\nu^2) = \nu$ und $\text{Var}(\chi_\nu^2) = 2\nu$ ([Fahrmeir et al., 2007](#)).

Somit entstehen bei Gleichsetzung der Momente

$$\begin{aligned}\nu &= \sigma^2 \text{Spur}(\mathbf{R}\mathbf{P}) \quad \text{und} \\ 2\nu &= 2\sigma^4 \text{Spur}((\mathbf{R}\mathbf{P})(\mathbf{R}\mathbf{P})).\end{aligned}$$

Beides nach σ^2 beziehungsweise σ^4 aufgelöst, ergibt

$$\sigma^2 = \frac{\nu}{\text{Spur}(\mathbf{R}\mathbf{P})} \tag{4.4}$$

$$\sigma^4 = \frac{\nu}{\text{Spur}((\mathbf{R}\mathbf{P})(\mathbf{R}\mathbf{P}))}. \tag{4.5}$$

Mit (4.4) in (4.5) kommt man schließlich zu

$$\nu = \frac{(\text{Spur}(\mathbf{R}\mathbf{P}))^2}{\text{Spur}((\mathbf{R}\mathbf{P})(\mathbf{R}\mathbf{P}))}, \tag{4.6}$$

dem approximierten Freiheitsgrad der t -Verteilung für die Teststatistik (4.2).

Eine weitere Approximation entsteht in der Anwendung dadurch, dass in (4.6) die wahre Korrelationsstruktur \mathbf{P} der Residuen unbekannt ist und aus den Daten geschätzt werden muss (vgl. Abschnitt 4.3.1).

4.1.2. Pre-Whitening

Ein zweiter Ansatz im Umgang mit abhängigen Stichprobenelementen ist ein sogenanntes Pre-Whitening. Das bedeutet, dass die zeitliche Korrelation aus den Daten eliminiert wird, bevor dann auf die somit unkorrelierten Daten ein klassisches Testverfahren angewendet wird ([Friston et al., 2007](#)).

Die Abhängigkeit der Stichprobenelemente im linearen Modell (4.1) kann, wenn die Residuen wie hier einer Normalverteilung unterliegen, anhand einer Linksmultiplikation beeinflusst werden (vgl. Satz 2.2), da sich diese allein auf die Abhängigkeit der Zeilen auswirkt. Dadurch kann das zeitlich korrelierte Modell auf die Standardverteilung mit unabhängigen Zeilen zurückgeführt werden, indem alle Teile der Modellgleichung mit der inversen Wurzel der Abhängigkeitsstruktur \mathbf{P} multipliziert werden

$$\mathbf{y}^* = \mathbf{P}^{-\frac{1}{2}}\mathbf{y} = \mathbf{P}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}^{-\frac{1}{2}}\boldsymbol{\epsilon} = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^*, \quad \mathbf{P}^{-\frac{1}{2}}\boldsymbol{\epsilon} = \boldsymbol{\epsilon}^* \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n). \tag{4.7}$$

Ist \mathbf{P} bekannt, kann die über die Residuen vorhandene Korrelation auf diese Weise vollständig aus den Daten eliminiert werden. Dies ermöglicht exakte Testverfahren wie den klassischen t -Test (2.12), indem $\mathbf{X}^* = \mathbf{P}^{-\frac{1}{2}}\mathbf{X}$ und $\mathbf{y}^* = \mathbf{P}^{-\frac{1}{2}}\mathbf{y}$ anstelle von \mathbf{X} und \mathbf{y} genutzt werden.

Ist die Abhängigkeitsstruktur \mathbf{P} jedoch unbekannt, muss sie geschätzt werden, so dass ein anschließender Test lediglich approximativ ist (vgl. Abschnitt 4.3.1).

4.2. Analoge multivariate Anpassungsstrategien für zeitlich korrelierte Daten

4.2.1. Satterthwaite-Approximation

In der Approximationsstrategie, wie sie im univariaten Modell zum Einsatz kommt (vgl. Abschnitt 4.1.1), wird die Korrelation der Daten über die Varianz des hypothesenbezogenen Kontrasts in die t -Teststatistik eingearbeitet und der Freiheitsgrad der unterliegenden Verteilung entsprechend approximiert. Diese Strategie soll analog im multivariaten Fall umgesetzt werden, was bereits in Kose (2007) Anwendung findet.

Bei der erweiterten Verteilungsannahme basieren die in dieser Arbeit hauptsächlich verwendeten Testverfahren auf der Summe der Produktsummenmatrizen \mathbf{G} und \mathbf{H} nach (2.20) und (2.19), wobei \mathbf{H} die hypothesenbezogenen Abweichungen charakterisiert. Ein Einbinden der Korrelationsstruktur \mathbf{P} ist hier über

$$\tilde{\mathbf{H}} = \mathbf{Y}'\mathbf{Q}_{\tilde{\mathbf{H}}}\mathbf{Y}$$

mit

$$\mathbf{Q}_{\tilde{\mathbf{H}}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}\left(\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}\right)^{-1}\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (4.8)$$

möglich (vgl. Kose, 2007). Die totale Produktsummenmatrix wird dann anhand $\tilde{\mathbf{W}} = \tilde{\mathbf{H}} + \mathbf{G}$ gebildet.

Bei unabhängigen Stichprobenelementen ($\mathbf{P} = \mathbf{I}$) lässt sich $\mathbf{Q}_{\tilde{\mathbf{H}}}$ wieder auf die originale Projektionsmatrix $\mathbf{Q}_{\mathbf{H}}$ zurückführen. In jedem Fall ist $\mathbf{Q}_{\tilde{\mathbf{H}}}$ weiterhin symmetrisch und es gilt ebenso $\mathbf{Q}_{\tilde{\mathbf{H}}}\mathbf{Q}_{\mathbf{G}} = \mathbf{0}$. Allerdings geht durch die Einarbeitung der Korrelation die Eigenschaft der Idempotenz in (4.8) verloren.

Gehen die veränderten Produktsummenmatrizen nun in eine multivariate Teststatistik wie Wilks' Λ ein, muss auch hier der mit dem Testdesign assoziierte Freiheitsgrad der Verteilung der veränderten Varianzstruktur angepasst werden. In Kose (2007) wie auch hier im Weiteren geschieht dies analog zum univariaten Fall anhand der Approximation (4.6).

4.2.2. Pre-Whitening

Das Pre-Whitening-Verfahren ist im multivariaten Modell in gleicher Weise durchführbar wie im univariaten. So gilt analog zu (4.7)

$$\mathbf{Y}^* = \mathbf{P}^{-\frac{1}{2}}\mathbf{Y} = \mathbf{P}^{-\frac{1}{2}}\mathbf{X}\mathbf{B} + \mathbf{P}^{-\frac{1}{2}}\mathbf{E} = \mathbf{X}^*\mathbf{B} + \mathbf{E}^*, \quad \mathbf{E}^* \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n \otimes \mathbf{\Sigma}).$$

Denn auch multivariat wirkt sich die Linksmultiplikation mit der inversen Wurzel der Abhängigkeitsstruktur nur innerhalb der Variablen und nicht auf die Kovarianzstruktur der Variablen untereinander aus (vgl. Satz 2.2 und Abschnitt 4.1.2). Somit kann auch hier die durch die Verteilung der Residuen bedingte Korrelation innerhalb der Stichprobenvektoren eliminiert und ein klassisches oder stabilisiertes multivariates Testverfahren angewendet werden.

4.3. Überlegungen zur Umsetzung der parametrischen Adaptionstrategien

Die in den Unterkapiteln 4.1 und 4.2 vorgestellten Adaptionstrategien der Testverfahren für die zeitliche Korrelation der Messungen basieren jeweils auf der Annahme einer bestimmten Korrelationsstruktur im Rahmen des linearen Modells.

Zweifelsfrei handelt es sich bei fMRT-Daten aufgrund der Signalveränderung im Gehirn bei Aktivierung um korrelierte Messungen, selbst wenn diese in Abständen von einigen Sekunden gemacht werden (vgl. Abschnitt 3.2.3). Die Abhängigkeitsstruktur entspricht dabei einem Prozess, bei dem die positive Korrelation der Messungen abnimmt, je länger diese voneinander entfernt aufgenommen werden. Eine konkrete Annahme über diese Struktur ist jedoch für die parametrische Anpassung der Testverfahren unumgänglich. Für die funktionelle Bildgebung wird dabei in der Regel ein autoregressiver Prozess erster Ordnung (Friston *et al.*, 2007) angenommen.

4.3.1. Schätzung der Korrelationsstruktur

In der praktischen Anwendung ist die zeitliche Korrelationsstruktur der Daten unbekannt und muss für beide bisher vorgestellten Adaptionstrategien geschätzt werden.

Unter der speziellen Annahme, dass die Verteilung der Daten allein der der theoretischen Residuen entspricht, also $\mathbf{Y} \sim N(\mathbf{0}, \mathbf{P} \otimes \mathbf{\Sigma})$, kann für den Verlauf der Messungen ein autoregressiver Prozess erster Ordnung unterstellt und dessen Korrelationskoeffizient ρ (vgl. Abschnitt 3.2.3) anhand der Daten geschätzt werden. Auf diese Schätzung soll im Weiteren genauer eingegangen werden.

Die Autokorrelation einer Messreihe $\mathbf{y} = (y_1, \dots, y_n)'$ kann geschätzt werden, indem die Korrelation zwischen der Messreihe $(y_1, \dots, y_{n-1})'$ mit sich selbst verschoben um – im Falle des AR(1)-Prozesses – den Lag 1 $(y_2, \dots, y_n)'$ geschätzt wird (Schlittgen & Streitberg, 2001). Sie ergibt sich aus dem Quotienten der empirischen Kovarianz (zum Lag 1)

$\hat{\gamma}(1) = \frac{1}{n-2} \sum_{i=1}^{n-1} (y_i - \bar{y})(y_{i+1} - \bar{y})$ und der empirischen Varianz $\hat{\sigma}^2 = \hat{\gamma}(0) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$:

$$\hat{\rho} = \frac{\hat{\gamma}(1)}{\hat{\sigma}^2}.$$

In der Zeitreihenanalyse wird die Summe für die empirische Kovarianz (einheitlich für alle Lags) jedoch meist durch n dividiert. Das hat den statistischen Vorteil, dass eine direkte Verbindung zu den Yule-Walker-Gleichungen existiert. Denn fasst man die Gleichung eines autoregressiven Prozesses (auch höherer Ordnungen) als ein Regressionsmodell auf, entspricht das Gleichungssystem für die Schätzungen der Regressionsparameter durch Einsetzen von $n\hat{\gamma}(\tau)$ für sämtliche Lags $\tau \geq 0$ den Yule-Walker-Gleichungen (Schlittgen & Streitberg, 2001; Kapitel 6). Somit ergibt sich die etablierte Schätzung des AR(1)-Korrelationskoeffizienten, wie sie beispielsweise auch in der SAS-Prozedur für Regression

implementiert ist (SAS, 2008),

$$\hat{\rho} = \frac{\sum_{i=1}^{n-1} (y_i - \bar{y})(y_{i+1} - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (4.9)$$

Die Verfälschung des Schätzers ist in den typischen Anwendungen häufig gering, da sich für größere n keine wesentlichen Unterschiede in der asymptotischen Verteilung ergeben (Schlittgen & Streitberg, 2001). Die Schätzung des Korrelationskoeffizienten gilt jedoch als nicht erwartungstreu (Schlittgen & Streitberg, 2001; SAS, 2008). Schlittgen & Streitberg (2001) geben den Erwartungswert für die Korrelationsschätzung für normalverteilte Messwerte für einen autoregressiven Prozess der Ordnung a an. Dieser kann für den AR(1)-Prozess vereinfacht werden zu

$$E(\hat{\rho}) = \rho - \frac{1}{n} [1 + 4\rho] + O(n^{-2}) \quad (4.10)$$

(siehe Anhang B).

Ersetzt man $E(\hat{\rho})$ in (4.10) mit $\hat{\rho}$ und vernachlässigt $O(n^{-2})$, so erhält man eine für den Erwartungswert bedingte Korrektur der Koeffizientenschätzung in Form von

$$\hat{\rho}_{\text{korrr.}} = \frac{n\hat{\rho} + 1}{n - 4}, \quad n > 4. \quad (4.11)$$

Im allgemeinen Fall sind im Modell (3.3) $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$ aber nur die theoretischen Residuen mit $\mathbf{E} \sim N(\mathbf{0}, \mathbf{P} \otimes \boldsymbol{\Sigma})$ verteilt, in deren Kovarianzstruktur die zeitliche Korrelation \mathbf{P} als ein autoregressiver Prozess erster Ordnung angenommen wird. Die Struktur der Daten entspricht jedoch nicht allein der der theoretischen Residuen. Sie unterliegt zusätzlich dem Einfluss der Modellparameter. Die Schätzung der Korrelationsstruktur muss daher anhand der geschätzten Residuen erfolgen. Diese werden erzeugt, indem die potentiellen Effekte anhand einer Linksmultiplikation der residuenbildenden Matrix \mathbf{R} an die Daten

$$\check{\mathbf{Y}} = \mathbf{R}\mathbf{Y}, \quad \mathbf{R} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (4.12)$$

aus diesen eliminiert werden (vgl. Abschnitt 2.2). Die Bereinigung der Daten um die Regressoreneffekte führt jedoch zu einer Veränderung der Korrelationsstruktur. Für die geschätzten Residuen $\check{\mathbf{Y}}$ ergibt sich die Kovarianzstruktur der Form $(\mathbf{R}\mathbf{P}\mathbf{R}') \otimes \boldsymbol{\Sigma}$, die somit für die Stichprobenvektoren nicht mehr einem reinen autoregressiven Prozess \mathbf{P} entspricht. Prinzipiell müsste demnach jeweils ein Modell der Korrelationsstruktur in Abhängigkeit des fMRT-Paradigmas erstellt und geschätzt werden. Dies ist jedoch aufwändig und wird mitunter vernachlässigt, so dass die zuvor beschriebene Schätzung der Korrelation im Sinne des Schätzens des AR(1)-Korrelationskoeffizienten trotz veränderter Abhängigkeitsstruktur anhand der geschätzten Residuen durchgeführt wird.

Eine direkte Umsetzung dieses Schätzproblems bieten Bayes'sche Verfahren, die in dieser Arbeit über die Prozedur MCMC im Programm SAS (SAS, 2008) realisiert werden.

Hierbei wird eine Maximum-Likelihood-Schätzung der Modellparameter iterativ über ein Markov-Chain-Monte-Carlo-Verfahren bestimmt. Für einen n -dimensionalen Datenvektor \mathbf{y} liegt dabei das Modell (4.1) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{P})$ sowie die multivariate Normalverteilung von \mathbf{y} (vgl. Definition 2.1) zugrunde.

Mit Hilfe der Ausdrücke für

$$|\sigma^2 \mathbf{P}| = \sigma^{2n} (1 - \rho^2)^{n-1} \quad \text{und} \quad (\sigma^2 \mathbf{P})^{-1} = \frac{1}{\sigma^2 (1 - \rho^2)} \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1 + \rho^2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 + \rho^2 & -\rho \\ 0 & \cdots & 0 & -\rho & 1 \end{pmatrix}$$

aus Lindsey (1994) sowie Klein & Zezula (2009) kann die Log-Likelihood-Funktion der Verteilungsparameter aufgeschrieben werden:

$$l(\boldsymbol{\beta}, \sigma^2, \rho | \mathbf{y}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{n-1}{2} \ln(1 - \rho^2) - \frac{1}{2\sigma^2(1 - \rho^2)} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1 + \rho^2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 + \rho^2 & -\rho \\ 0 & \cdots & 0 & -\rho & 1 \end{pmatrix} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (4.13)$$

Anhand von A-priori-Annahmen über die Verteilung der einzelnen Parameter wird diese Funktion über Markov-Chain-Monte-Carlo-Iterationen maximiert. Die so erzielten A-posteriori-Schätzungen stellen die optimierten Parameterwerte für die unterliegenden Daten \mathbf{y} entsprechend der Modellannahmen (4.1) dar. Die resultierende Schätzung des AR(1)-Korrelationskoeffizienten ist somit nicht durch die Residuenbildung verzerrt.

Da speziell in den parametrischen Anpassungsverfahren für die zeitliche Korrelation der Daten die adäquate Schätzung der zugrunde liegenden Abhängigkeitsstruktur von essentieller Bedeutung ist, sollte eine bestmögliche Schätzung des Korrelationskoeffizienten der angenommenen AR(1)-Struktur erfolgen. Die verschiedenen Schätzmöglichkeiten des AR(1)-Korrelationskoeffizienten werden daher im Weiteren anhand simulierter Daten einander gegenübergestellt. Analog zu den weiteren Simulationen zur Untersuchung der Test- und Adaptionsverfahren in den Kapiteln 5 und 6 werden $n = 250$ fMRT-Aufnahmen simuliert. Aus zeit- und programmiertechnischen Gründen geschieht dies hier jedoch nur als Betrachtung einer Variable ($p = 1$). Eine Einarbeitung räumlicher Korrelationsstrukturen bleibt damit aus. Die Messungen werden unter der globalen Nullhypothese, dass keine der Kovariablen einen Einfluss hat, und mit zugrunde liegender AR(1)-Korrelationsstruktur für verschiedene Korrelationskoeffizienten $\rho \in \{0, 0; 0, 2; 0, 4; 0, 6; 0, 8\}$ simuliert. Dabei werden beide Formen des fMRT-Paradigmas, das in die Designmatrix des Modells einfließt, simuliert: ein Blockdesign und ein ereignisbezogenes Design (vgl. Abschnitt 3.2.2). Weitere Details zur Simulation der Daten

sind im Abschnitt 6.1 aufgeführt.

Die anhand der Residuen bestimmte konventionelle Schätzung (4.9) sowie ihre erwartungswertkorrigierte Version (4.11) werden einer Bayes-Schätzung des AR(1)-Korrelationskoeffizienten gegenübergestellt. Für die Bayes-Schätzung werden als A-priori-Verteilungen der Parameter folgende Annahmen getroffen:

$$\begin{aligned}\beta_j &\sim U(-50, 50) \quad \forall j, \\ \sigma &\sim \Gamma(3, 3), \\ \rho &\sim U(0, 1),\end{aligned}$$

mit $j = 3$ Kovariablen beim Blockdesign und $j = 4$ beim ereignisbezogenen Design (vgl. Abschnitt 6.1). Als A-priori-Verteilung für den Korrelationskoeffizienten wird hierbei eine Gleichverteilung auf $[0, 1]$ angenommen. Damit geht zwar etwas Vorwissen über den zu schätzenden Parameter in die A-priori-Verteilung ein (Beschränkung auf positive Werte), genaueren Annahmen über die jeweils simulierte Korrelationsstärke werden aber nicht getroffen.

Die Simulationsanzahl für die Schätzungen beträgt $n_{sim} = 2500$, wobei alle Schätzungen an denselben simulierten Daten vorgenommen und innerhalb jeder Simulation für die Bayes-Schätzung 50 000 MCMC-Iterationen durchgeführt werden.

Design	Schätzung	ρ				
		0,0	0,2	0,4	0,6	0,8
Block	$\hat{\rho}$ (4.9)	-0,01084	0,18432	0,37923	0,57647	0,77336
	$\hat{\rho}_{\text{korr.}}$ (4.11)	-0,00695	0,19138	0,38947	0,58991	0,79000
	$\hat{\rho}_{\text{Bayes}}$ U(0,1)	0,05923	0,20333	0,40180	0,60445	0,80618
Event	$\hat{\rho}$ (4.9)	-0,01041	0,18470	0,37965	0,57618	0,77127
	$\hat{\rho}_{\text{korr.}}$ (4.11)	-0,00652	0,19177	0,38989	0,58962	0,78788
	$\hat{\rho}_{\text{Bayes}}$ U(0,1)	0,05911	0,20320	0,40169	0,60380	0,80422

Tabelle 4.1.: Mittelwerte der Schätzungen des AR(1)-Korrelationskoeffizienten für simulierte fMRT-Daten und beide fMRT-Designs ($n = 250$, 50 000 MCMC-Iterationen, 2500 Wiederholungen)

In Tabelle 4.1 sind für den simulierten fMRT-Datenvektor die Mittelwerte der Schätzungen des AR(1)-Korrelationskoeffizienten aus je 2500 Simulationen für beide fMRT-Paradigmentypen angegeben. Dabei steht Block für das Blockdesign und Event für das ereignisbezogene Design.

Weiterhin wird in Tabelle 4.2 der Root Mean Squared Error

$$\text{RMSE}_{\hat{\rho}} = \sqrt{\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} (\hat{\rho}_i - \rho)^2} \quad (4.14)$$

für die verschiedenen Schätzverfahren angegeben. Der $\text{RMSE}_{\hat{\rho}}$ dokumentiert die Abweichungen der einzelnen Schätzungen $\hat{\rho}_i$ vom wahren Wert ρ für die $n_{sim} = 2500$ Simulationen.

		ρ				
Design	RMSE der Schätzung	0,0	0,2	0,4	0,6	0,8
Block	$\text{RMSE}_{\hat{\rho}}$ (4.9)	0,06386	0,06350	0,06156	0,05636	0,04926
	$\text{RMSE}_{\hat{\rho}_{\text{korr.}}}$ (4.11)	0,06433	0,06313	0,05983	0,05301	0,04328
	$\text{RMSE}_{\hat{\rho}_{\text{Bayes}}}$ U(0,1)	0,06548	0,06017	0,05960	0,05235	0,04237
Event	$\text{RMSE}_{\hat{\rho}}$ (4.9)	0,06404	0,06353	0,06139	0,05637	0,05050
	$\text{RMSE}_{\hat{\rho}_{\text{korr.}}}$ (4.11)	0,06455	0,06320	0,05972	0,05295	0,04392
	$\text{RMSE}_{\hat{\rho}_{\text{Bayes}}}$ U(0,1)	0,06536	0,06025	0,05948	0,05188	0,04026

Tabelle 4.2.: $\text{RMSE}_{\hat{\rho}}$ für die Schätzung des AR(1)-Korrelationskoeffizienten für simulierte fMRT-Daten und beide fMRT-Designs ($n = 250$, 50 000 MCMC-Iterationen, 2500 Wiederholungen)

Im Vergleich der Schätzungen erweist sich das Bayes-Verfahren basierend auf der Gleichverteilung auf $[0, 1]$ als genaueste Methode. Abgesehen von der Schätzung für $\rho = 0, 0$ am Rand der angegebenen A-priori-Verteilung ergeben sich hier generell die besten Schätzer und die kleinsten Werte für den Root Mean Squared Error. Dabei liegen die mittleren Schätzungen für beide Designs jeweils leicht über dem wahren Wert, während die Schätzungen basierend auf den empirischen Residuen diesen eher unterschätzen. Erwartungsgemäß liefert die erwartungswertkorrigierte Schätzung eine Verbesserung des konventionellen Schätzers.

Beispielsweise kann für das Blockdesign bei einer wahren Autokorrelation von $\rho = 0, 4$ der mittlere Schätzwert von $\hat{\rho} = 0, 37923$ auf $\hat{\rho}_{\text{korr.}} = 0, 38947$ verbessert werden, wobei sich die entsprechenden Abweichungen der Einzelwerte von $\text{RMSE}_{\hat{\rho}} = 0, 06156$ auf $\text{RMSE}_{\hat{\rho}_{\text{korr.}}} = 0, 05983$ verringern. Obwohl die mittlere Schätzung durch das Bayes-Verfahren in diesem Fall mit $\hat{\rho}_{\text{Bayes}} = 0, 40180$ deutlich näher am wahren Wert ist, sind die Abweichungen der einzelnen Werte mit $\text{RMSE}_{\hat{\rho}_{\text{Bayes}}} = 0, 05960$ nur geringfügig besser als die der korrigierten Schätzung anhand der empirischen Residuen.

Die Schätzung durch das Markov-Chain-Monte-Carlo-Verfahren ist jedoch stark von der A-priori-Verteilungsannahme abhängig. Tabelle 4.3 stellt für beide fMRT-Designs und einen ausgewählten AR(1)-Korrelationskoeffizienten ($\rho = 0, 4$) dar, wie stark die Schätzungen durch Annahme anderer A-priori-Verteilungen für ρ beeinflusst werden (die Annahmen über die anderen Parameterverteilungen bleiben unverändert). Auch hier werden die mittlere Schätzung des Koeffizienten anhand 2500 Wiederholungen und der entsprechende Root Mean Squared Error nach (4.14) angegeben, wobei jede Schätzung durch 50 000 MCMC-Iterationen gebildet wird.

Wird als A-priori-Verteilung von ρ eine Gleichverteilung über den gesamten oder positiven Wertebereich des Korrelationskoeffizienten angenommen, unterscheiden sich die bei-

		A-priori-Verteilung				
Design	$\rho = 0,4$	U(-1, 1)	U(0, 1)	N(0,2, 0,1)	N(0,4, 0,1)	N(0,6, 0,1)
Block	$\hat{\rho}_{\text{Bayes}}$	0,40181	0,40180	0,34806	0,40179	0,45606
	RMSE $_{\hat{\rho}_{\text{Bayes}}}$	0,05958	0,05959	0,06908	0,04372	0,06975
Event	$\hat{\rho}_{\text{Bayes}}$	0,40158	0,40169	0,34822	0,40173	0,45540
	RMSE $_{\hat{\rho}_{\text{Bayes}}}$	0,05931	0,05948	0,06897	0,04358	0,06913

Tabelle 4.3.: Mittelwerte der Bayes-Schätzungen des AR(1)-Korrelationskoeffizienten $\rho = 0,4$ und RMSE der Schätzungen für verschiedene A-priori-Verteilungen ($n = 250$, 50 000 MCMC-Iterationen, 2500 Wiederholungen)

den mittleren Schätzungen sowohl beim Block- als auch beim ereignisbezogenen Paradigma kaum. Die Abweichungen fallen mit $\text{RMSE}_{\hat{\rho}_{\text{Bayes}}} = 0,05931$ bis $\text{RMSE}_{\hat{\rho}_{\text{Bayes}}} = 0,05959$ relativ gering aus. Bessere Ergebnisse werden nur erzielt, wenn eine detailliertere Information in die A-priori-Verteilungsannahme einfließt. Für die wahre Korrelationsstärke von $\rho = 0,4$ weist die Schätzung auf Basis einer Normalverteilungsannahme mit einem wahren Mittel von 0,4 und einer Standardabweichung von 0,1 die beste Schätzung auf ($\text{RMSE}_{\hat{\rho}_{\text{Bayes}}} = 0,04372$ beim Blockdesign beziehungsweise $\text{RMSE}_{\hat{\rho}_{\text{Bayes}}} = 0,04358$ beim ereignisbezogenen Design). Weicht die A-priori-Verteilungsannahme jedoch zu stark vom wahren Wert ab, wie hier durch die beiden weiteren Normalverteilungsannahmen simuliert, wird die A-posteriori-Schätzung verzerrt.

Anhand eines Bayes-Verfahrens kann eine sehr gute Schätzung der wahren Korrelation realisiert werden, wenn die A-priori-Annahme für den Korrelationsschätzer adäquat ist. In der eigentlichen Anwendung sind die Vorkenntnisse jedoch eher gering (etwa der Annahme U(0,1) entsprechend), da sie einem bestimmten Kontext (Gehirnregion, Anzahl der Modellparameter, Datenvorbehandlung, usw.) unterliegen. Zudem ist diese Schätzmethode im Vergleich zu denen, die auf den empirischen Residuen basieren, sehr aufwändig: Die Log-Likelihood-Funktion muss aufgestellt werden und es sind zur iterativen Optimierung der Parameterschätzer zahlreiche rechenintensive MCMC-Schritte notwendig. Die konventionelle Schätzung ist hingegen durch eine einzige Berechnung anhand der empirischen Residuen erzielbar, wobei der Schätzer durch die Erwartungswertkorrektur deutlich verbessert werden kann.

Aufgrund des enormen Aufwandes – gerade innerhalb von Simulationsuntersuchungen – bei einer relativ geringen Verbesserung der Schätzung, wird in dieser Arbeit im Weiteren die erwartungswertkorrigierte residuenbasierte Schätzung des AR(1)-Korrelationskoeffizienten genutzt. Für sämtliche parametrischen Analysen von simulierten und realen Daten wird einheitlich eine zeitliche AR(1)-Abhängigkeitsstruktur innerhalb aller Voxel angenommen (vgl. Friston *et al.*, 2007, S. 146), der Korrelationskoeffizient je Voxel nach (4.9) geschätzt und nach (4.11) korrigiert. Anschließend wird diese Schätzung über alle p Voxel gemittelt (vgl. Abschnitt 6.1).

4.3.2. Andere parametrische Ansätze zur Schätzung der Kovarianzstruktur

Die in dieser Arbeit für die Residuen verwendete Kovarianzstruktur, insbesondere die Darstellung der Struktur als Kronecker-Produkt aus räumlichen und zeitlichen Abhängigkeiten, ist in verschiedenen Arbeiten zum Thema Raum-Zeit-Struktur zu finden. Dabei wird die Schätzung auf verschiedene Weise realisiert.

Beispielsweise nutzen [Huizenga et al. \(2002\)](#) diese Verteilungsstruktur für die Untersuchung von EEG/MEG-Daten. Die Schätzung der Kovarianz wird durch die Formulierung als Kronecker-Produkt zweier separater Kovarianzmatrizen zur getrennten Schätzung der Raum-Zeit-Kovarianzen und somit vereinfacht. Dabei wird auch hier für die zeitlichen Abhängigkeiten eine autoregressive Struktur angenommen, deren Parameter mittels Maximum-Likelihood-Funktionen geschätzt wird.

[Ohlson & Koski \(2009\)](#) nutzen die matrixvariante Normalverteilung mit separaten Kovarianzmatrizen zur Bildung einer Likelihood-Ratio-Teststatistik zum Test auf räumliche Unkorreliertheit. Unter Annahme eines AR(1)-Prozesses für die zeitliche Abhängigkeitsstruktur und der möglichen Ausnutzung der daraus einfach herzuleitenden Determinante und Inversen (vgl. Abschnitt 4.3.1 sowie [Lindsey, 1994](#); [Klein & Zezula, 2009](#)), schlagen sie zwei Algorithmen zur effektiven Schätzung der räumlichen Kovarianz vor. Innerhalb der Algorithmen wird darüber hinaus die zunächst verwendete konventionelle Schätzung (4.9) des AR(1)-Korrelationskoeffizienten optimiert.

[Wang & West \(2009\)](#) erweitern einen Bayes'schen Ansatz für grafische Modelle auf Basis der matrixvarianten Normalverteilung mit räumlich und zeitlich separierter Kovarianzstruktur. Sie nutzen dabei die sogenannten Präzisionsmatrizen, die den inversen Kovarianzmatrizen und somit den partiellen Korrelationen entsprechen. Bei einem reinen AR(1)-Prozess ist die partielle Korrelation zum Lag τ , $\tau > 1$, stets Null ([Schlittgen & Streitberg, 2001](#)). [Wang & West](#) nehmen jedoch keine feste Ordnung des Prozesses an, so dass innerhalb der Nebendiagonalen partielle Korrelationen ungleich Null auftreten können, die entsprechende Präzisionsmatrix aber sehr dünn besetzt ist. Auf diese Struktur werden nun Graphen angewendet, die über Likelihood-Schätzungen optimiert werden.

Die Schätzproblematik der Kovarianzstruktur für multivariate Daten wird also durch Verwendung der Kronecker-Struktur mit zwei separaten Raum-Zeit-Komponenten vereinfacht. In den zitierten Arbeiten müssen teilweise aber Stichproben von mehreren unabhängigen Raum-Zeit-Realisierungen vorliegen. Im Kontext der fMRT-Untersuchungen entspräche dies der Betrachtung mehrerer Probanden, was üblicherweise in sogenannten second-level-Analysen umgesetzt wird. In diese Analysen fließen jedoch nicht die gesamten Messungen ein, sondern die einer Nullhypothese entsprechenden Kontraste. Dies soll aber in dieser Arbeit nicht weiter betrachtet werden, da hier der Fokus auf verschiedenen Ansätzen der Datenanalyse einer einzelnen Probandensitzung liegt, in welchen die Raum-Zeit-Kovarianzstruktur inbegriffen ist. Für die hier untersuchten funktionel-

len Bilddaten besteht also vor allem das Problem der optimalen Schätzung der zeitlichen Kovarianzstruktur.

Auch in der funktionellen Bildgebung finden Bayes'sche Ansätze zur Schätzung der zeitlichen Abhängigkeiten Anwendung ([Friston *et al.*, 2002](#); [Frackowiak *et al.*, 2003](#); [Penny *et al.*, 2007](#)). Als Annahme über die Korrelationsstruktur dient dabei in der Regel der autoregressive Prozess erster Ordnung. Möglicherweise ist aber bereits diese Annahme in Abhängigkeit der untersuchten Hirnregion, des Paradigmas oder anderer Confounder nicht immer zutreffend.

Beispielsweise könnte aufgrund der relativ begrenzten Dauer der Signaländerung nach einem kurzen Reiz (vgl. hämodynamische Responsefunktion (3.2)), das Abklingen eines AR(1)-Prozesses die Korrelationen bei höheren Lags überschätzen. In solch einem Fall wäre eine Bandmatrixstruktur ([Gray, 2005](#)) realistischer, deren Haupt- und ersten r Nebendiagonalen einer AR(1)-Struktur entsprechen, die restlichen $n - 1 - r$ Nebendiagonalen jedoch Nullen enthalten. Dies hätte auch Vorteile für das im folgenden Kapitel vorgestellte nichtparametrische Verfahren zum Umgang mit der zeitlichen Abhängigkeit. Andererseits ist bei Modellierungen der Signalveränderung unter verschiedenen Bedingungen, wie beispielweise Alkoholeinfluss, sichtbar, dass der Verlauf der Signaländerung stark variieren kann ([Luchtmann, 2010](#)). Eine gleichmäßig mit steigender Potenz abklingende Korrelation wie bei einem AR(1)-Prozess würde demnach den Signalverlauf nur unzureichend widerspiegeln. In solch einem Fall wäre eine unstrukturierte Korrelationsannahme, wie sie etwa auch [Ahmad \(2008\)](#) verwendet, zutreffender, bei der sämtliche Korrelationskoeffizienten einzeln nur in Abhängigkeit des Lags τ geschätzt werden müssten. Dies würde jedoch den Aufwand, vor allem im Vergleich zu einem AR(1)-Prozess, bei dem nur ein Parameter geschätzt werden muss, erheblich erhöhen. Zudem birgt die Schätzung vieler Parameter größere Gefahr für Schätzfehler, weshalb sich – gerade bei geringen Stichprobenumfängen – sparsame, wenn auch approximativere Modellierungsansätze als günstiger erwiesen haben.

Letztlich muss unter Nutzung parametrischer Verfahren zur Auswertung von funktionellen Bilddaten grundsätzlich eine Annahme über die Korrelationsstruktur der Stichprobenelemente beziehungsweise -vektoren getroffen werden.

Ob dies – wie im Großteil der fMRT-Analysen realisiert – über einen autoregressiven Prozess erster Ordnung optimal ist, soll hier nicht beantwortet werden. Doch selbst unter dieser Annahme wird im vorausgehenden Kapitel gezeigt, dass die Schätzung des entsprechenden Korrelationskoeffizienten aus den Daten mit Schwierigkeiten behaftet ist. Aus diesem Grund wird im folgenden Kapitel eine nichtmetrische Adaptionvariante für zeitliche Korrelationen vorgestellt. Dabei handelt es sich um eine Anpassung von Permutationsverfahren, die dann im Weiteren den bisherigen parametrischen Überlegungen gegenübergestellt wird.

5. Adaption von Permutationsverfahren für den Umgang mit korrelierten Daten

Bei klassischen Permutationsverfahren beruht die Verteilungsannahme auf der Austauschbarkeit der Stichprobenelemente unter der Nullhypothese – und damit ihrer Unabhängigkeit (vgl. Abschnitt 2.3.3).

Bei Verwendung des in dieser Arbeit grundlegenden multivariaten linearen Modells (3.3) mit der theoretischen Residuenverteilung $\mathbf{E} \sim N_{n \times p}(\mathbf{0}, \mathbf{P} \otimes \mathbf{\Sigma})$ ist diese Austauschbarkeit unter der Nullhypothese aufgrund der zeitlichen Abhängigkeitsstruktur \mathbf{P} nicht mehr gegeben.

Bei funktionellen MRT-Scans korrelieren aufeinanderfolgende Messungen höher miteinander als solche, die weiter voneinander entfernt liegen. Für \mathbf{P} wird daher für fMRT-Daten in der Regel ein autoregressiver Prozess erster Ordnung angenommen (vgl. Abschnitt 3.2.3). Analog zu den parametrischen Verfahren muss eine Anpassung der Permutationsstrategie auf diese Datenstruktur erfolgen, um das nominale Testniveau einzuhalten.

Um einen Permutationstest zur Auswertung zeitlich abhängiger Daten nutzen zu können, wird in Abschnitt 5.1 eine blockweise Permutationsstrategie verfolgt. Hierbei werden nicht die einzelnen Stichprobenelemente, sondern ganze Blöcke aufeinanderfolgender Elemente permutiert. Um die dabei auftretenden Probleme bei kleinen Stichprobenumfängen abzufangen, wird in Abschnitt 5.2 eine Variation des blockweisen Permutierens anhand einer zufälligen Verschiebung vorgeschlagen.

5.1. Blockpermutationen

Ein Ansatz zur Berücksichtigung der Korrelation der Stichprobenelemente beruht darauf, ganze Blöcke benachbarter Elemente zu permutieren. Analog zu der klassischen Permutation bezieht sich dies im multivariaten Fall auf Blöcke zeitlich aufeinanderfolgender Vektoren.

Um das Verfahren der Blockpermutation selbst genauer zu erläutern, wird im Weiteren zunächst ein univariates Modell mit austauschbaren Beobachtungen unter H_0

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

mit der globalen Nullhypothese $H_0 : \boldsymbol{\beta} = \mathbf{0}$ betrachtet, so dass dessen Beobachtungen für die Durchführung eines Permutationstests permutiert werden können.

Im Fall eines Zweistichprobenproblems entspräche eine klassische elementweise Permutation der Darstellung in Abbildung 5.1. Dabei wird zunächst die Teststatistik der Originalstichprobe ermittelt. Anschließend wird die bedingte Verteilung der Teststatistik unter der Nullhypothese durch die Durchführung aller N_π oder einer zufälligen Anzahl n_π von Permutationen bestimmt (vgl. Abschnitt 2.3.3).

Klassische Permutation

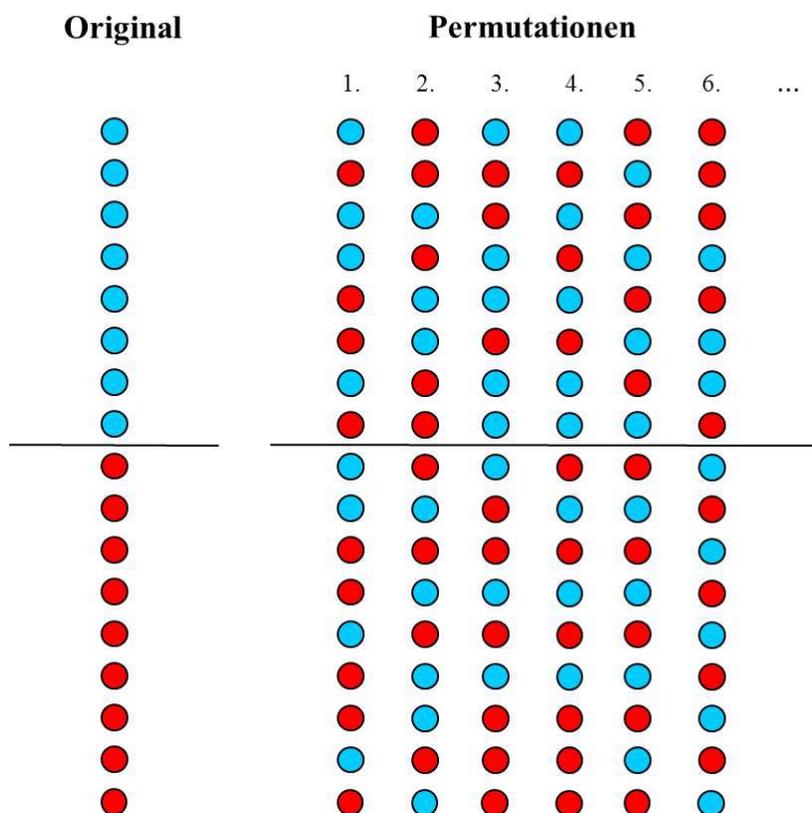


Abbildung 5.1.: Schematische Darstellung der Permutation eines univariaten Zweistichprobenproblems mit $n = 15$ Elementen

Bei der blockweisen Permutation werden nun statt einzelner Stichprobenelemente ganze Blöcke zeitlich benachbarter Beobachtungen gemeinsam betrachtet und in ihrer Ordnung verändert, wie in Abbildung 5.2 dargestellt.

Die Blöcke werden jeweils als ein Element betrachtet und permutiert. Die für die Permutation notwendige Bedingung der Austauschbarkeit der Elemente unter der Nullhypothese wird nunmehr zur Bedingung der Austauschbarkeit der Blöcke unter H_0 . Somit ergibt sich die Anzahl möglicher Permutationen direkt aus der Anzahl k zu permutierender Blöcke.

Blockweise Permutation

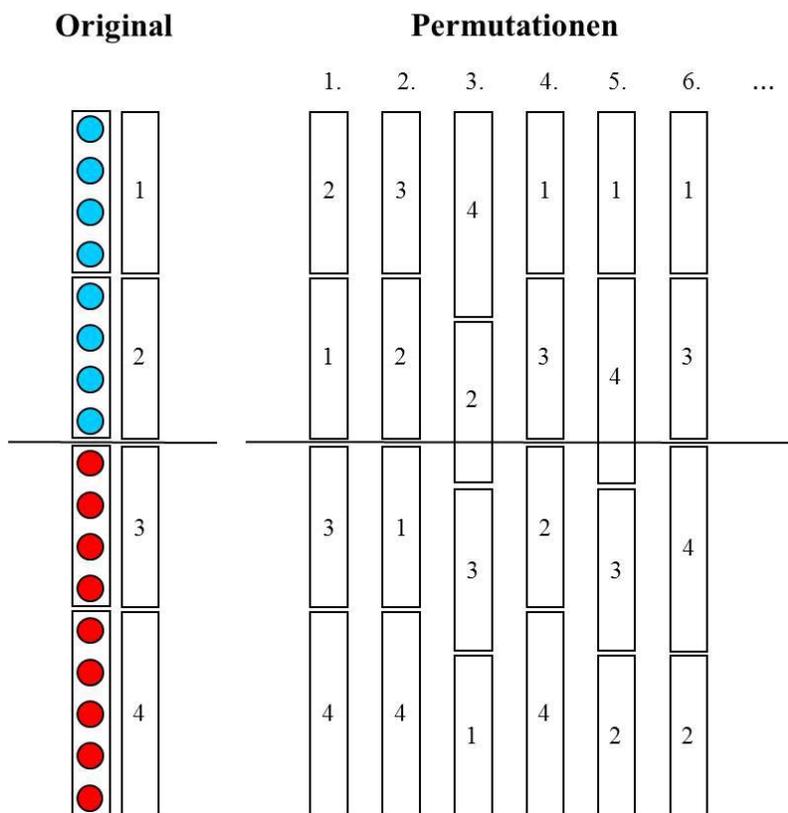


Abbildung 5.2.: Schematische Darstellung der blockweisen Permutation eines univariaten Zweistichprobenproblems mit Blocklänge $l = 4$ und Blockanzahl $k = 4$

Die Blockbildung erfolgt, indem der Beobachtungsvektor \mathbf{y} in aufeinanderfolgende Abschnitte der Länge l geteilt wird. Dabei liegt die Blocklänge l bei $1 \leq l \leq \frac{n}{2}$. Ist die Blocklänge l kein Teiler des Stichprobenumfangs n , $n \bmod l = r$, ($0 < r < l$), enthält der letzte Block $l+r$ Elementen beziehungsweise Vektoren.

Im Extremfall $l = 1$ entspricht dies der Einteilung in die einzelnen Stichprobenelemente, also $k = n$ Blöcken bestehend aus je einem Element, woraus sich eine klassische elementweise Permutation ergibt. Der Fall $l = \frac{n}{2}$ hingegen bewirkt eine Permutation mit nur zwei Blöcken.

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_l \\ y_{l+1} \\ \vdots \\ y_{2l} \\ y_{2l+1} \\ \vdots \\ y_{3l} \\ \vdots \\ \vdots \\ y_n \end{pmatrix} \quad (5.1)$$

Für nichtparametrische Testverfahren sind blockweise Strategien bereits bei einigen Anwendungen zu finden. Beispielsweise nutzen [Pesarin & Salmaso \(2010\)](#) eine andere Art blockweiser Permutation zur Reduzierung des rechentechnischen Aufwandes bei der Bildung der gesamten Permutationsverteilung für ein Testproblem. Sie transformieren ein eindimensionales Testproblem der Länge n in ein p -dimensionales Testproblem der Länge $\frac{n}{p}$. Die Austauschbarkeit der Blöcke unter der Nullhypothese wird dabei vorausgesetzt, die Kovarianzstruktur der entstandenen p Variablen ist aber zu vernachlässigen. Statt im Fall eines Zweistichprobenproblems insgesamt $\binom{n}{n_1}$ verschiedene univariate Permutationen durchzuführen, werden im transformierten Testproblem nur noch $\binom{n/p}{n_1/p}$ unterschiedliche multivariate Permutationen benötigt, so dass die Rechenzeit stark reduziert werden kann. Dabei benennen [Pesarin & Salmaso](#) mit dem Begriff der *equipower property* die Eigenschaft, dass mit Hilfe dieser Blockbildung nahezu äquivalente Ergebnisse in Form des Fehlers 1. und 2. Art entstehen. In ihren Simulationsuntersuchungen testen sie diesbezüglich neben standardnormalverteilten Daten noch weitere Verteilungen. Im Gegensatz zu der in dieser Arbeit verwendeten Blockbildung bestehen die Blöcke, die [Pesarin & Salmaso](#) permutieren, nicht aus aufeinanderfolgenden Stichprobenelementen. Aufgrund der Austauschbarkeit der einzelnen Elemente unter der Nullhypothese ist dies jedoch keine Restriktion.

Bei Unabhängigkeit der Stichprobenelemente beziehungsweise -vektoren ist eine blockweise Permutation also weitestgehend eine adäquate Alternative zur klassischen Permutationsstrategie. In dieser Arbeit soll die Blockbildung aufeinanderfolgender Messungen jedoch hauptsächlich dazu dienen, das nominale Testniveau trotz Korrelation der Messungen approximativ einzuhalten.

Dies ist möglich unter der speziellen Annahme, dass die zeitliche Abhängigkeitsstruktur der Daten einem autoregressiven Prozess erster Ordnung entspricht, oder aber allgemeiner betrachtet einem stationären Prozess, in dem die Korrelation der Messungen abnimmt, je weiter diese zeitlich auseinander liegen. Letzteres entspricht zweifellos der zeitlichen Struktur von fMRT-Daten.

Beispielsweise zeigt [Kirch \(2007\)](#) anhand eines AMOC-Prozesses, eines Prozesses mit abrupter Mittelwertänderung (Change Point) und abhängigen Fehlern, dass eine Blockpermutation für verschiedene Change-Point-sensitive Teststatistiken bei nach AR(1) abhängigen Fehlertermen in der Verteilungsasymptotik mit der originalen Teststatistik übereinstimmt. Dabei stellt sie anhand von Simulationsuntersuchungen dar, dass die optimale Blocklänge in Abhängigkeit von der Korrelationsstärke variiert.

Weiterhin kommt eine Blockbildung auch in Bootstrap-Methoden für korrelierte Daten zur Anwendung ([Efron, 1977](#); [Mignani & Rosa, 1995](#)): der sogenannte moving blocks bootstrap. Bei diesem speziellen Bootstrapping wird eine Korrelation innerhalb der Daten in Form eines stationären Prozesses berücksichtigt, indem nicht einfach einzelne Elemente zufällig ausgewählt und unter Zurücklegen zu einer Stichprobe kombiniert werden, sondern ganze Blöcke aufeinanderfolgender Elemente. Die Blockbildung erfolgt nach [Mignani & Rosa](#) jedoch fließend für alle möglichen Blöcke einer festgelegten Blocklänge l ,

also

$$\{y_1, y_2, \dots, y_l\}, \{y_2, y_3, \dots, y_{l+1}\}, \dots, \{y_{n-l+1}, \dots, y_{n-1}, y_n\}. \quad (5.2)$$

Damit entstehen insgesamt $n - l + 1$ Blöcke der Länge l , die unter Zurücklegen in die Bootstrap-Stichproben gewählt werden können. Die Intention dabei ist auch hier, mit einer hinreichenden Blocklänge nahezu eine Unabhängigkeit der in die Stichproben gezogenen Blöcke voneinander zu erreichen, während die Korrelationsstruktur innerhalb der Blöcke erhalten bleibt. In Abhängigkeit von der mit der Bootstrap-Prozedur zu beantwortenden Fragestellung ist auch in [Mignani & Rosa \(1995\)](#) die Festlegung der nötigen Blocklänge zwar von der Korrelationsstärke abhängig, lässt sich aber dennoch nicht generell optimal angeben.

Im Weiteren wird nun eine Blockpermutation für eine Zeitreihe funktioneller Aufnahmen betrachtet, wobei die Blöcke analog zur Darstellung (5.1) eine Blocklänge l haben. Bei Durchführung der Permutation können die Blöcke also als jeweils eine Einheit betrachtet werden. Diese sollten unter der Nullhypothese approximativ unabhängig voneinander und damit vertauschbar sein.

Bei Annahme eines autoregressiven Prozesses erster Ordnung mit dem Korrelationskoeffizienten ρ korreliert jedes Element eines beliebigen Blockes der Länge l $\{y_{i+1}, y_{i+2}, \dots, y_{i+l}\}$ mit dem entsprechenden Element des folgenden Blockes $\{y_{i+l+1}, \dots, y_{i+2l}\}$ anhand der Korrelationsstärke ρ^l . Weiterhin ergeben sich sämtliche elementweisen Korrelationen zweier aufeinanderfolgender Blöcke der Länge l aus

$$\begin{matrix} & y_{i+l+1} & y_{i+l+2} & \dots & y_{i+2l} \\ y_{i+1} & \left(\begin{array}{cccc} \rho^l & \rho^{l+1} & \dots & \rho^{2l-1} \\ \rho^{l-1} & \rho^l & \dots & \vdots \\ \vdots & \vdots & \dots & \rho^{l+1} \\ \rho^1 & \dots & \rho^{l-1} & \rho^l \end{array} \right) & = & (\rho_{st})_{l \times l}. \end{matrix} \quad (5.3)$$

Je weiter zwei Elemente aufeinanderfolgender Blöcke voneinander entfernt sind – also mit wachsender Blocklänge l – desto stärker sinkt ihre Korrelation. Man kann zwar durch die Blockbildung nie eine tatsächliche Unabhängigkeit der Blöcke zueinander erreichen, da die Randelemente aus aufeinanderfolgenden Blöcken stets die Korrelation ρ zueinander aufweisen. Doch mit größer werdender Blocklänge l treten diese hohen elementweisen Korrelationen der Randelemente in (5.3) im Verhältnis zu den übrigen Korrelationen weiter voneinander entfernter Elemente innerhalb zweier Blöcke in den Hintergrund. Das geometrische Mittel ([Sachs, 2004](#)) dieser elementweisen Korrelationen

$$\bar{\rho}_G = \sqrt[l]{\prod_{s,t=1}^l (\rho_{st})}$$

entspricht genau ρ^l , also der Korrelation eines Elements mit dem entsprechendem Element im angrenzenden Block, und verkleinert sich mit wachsender Blocklänge. Selbst bei hohen Korrelationskoeffizienten wie $\rho = 0,8$ gilt für das geometrische Mittel der Korrelation zweier benachbarter Blöcke $\bar{\rho}_G \leq 0,001$ bereits für eine Blocklänge von $l = 31$

und $\bar{\rho}_G \leq 0,0001$ gilt in diesem Fall ab $l = 42$.

Um eine annähernde Unabhängigkeit und damit Vertauschbarkeit der zu permutierenden Blöcke zu gewährleisten, muss also eine hinreichend große Blocklänge l gewählt werden, um das geometrische Mittel der Korrelationsstärken zwischen den Elementen zweier benachbarter Blöcke zu mindern.

Entspricht die Zeitreihe keinem autoregressiven Prozess erster Ordnung, zumindest aber einem stationären Prozess, bei dem die Korrelation der Elemente abnimmt, je weiter diese zeitlich auseinander liegen, ist es durchaus möglich, dass eine größere Blocklänge nötig ist, um annähernde Unabhängigkeit der Blöcke bei Permutation zu erreichen.

Eine große Blocklänge bedingt jedoch die Verkleinerung der Anzahl der Blöcke zur Bestimmung der Permutationsverteilung, also eine geringe Anzahl an verfügbaren Permutationsmöglichkeiten. Bei funktionellen Bilddaten sind diese Gegensätze nur begrenzt miteinander zu vereinbaren, da aufgrund der experimentellen Gegebenheiten meist nur eine Anzahl von einigen hundert Messungen zur Verfügung steht. Die Wahl einer großen Blocklänge verursacht somit eine geringe Verfügbarkeit unterschiedlicher Blöcke und damit eine Einschränkung der Permutationsmöglichkeit. Eine Lösung zum Umgang mit diesem Problem bietet die Einbeziehung einer zufälligen Verschiebung, die im folgenden Unterkapitel vorgestellt wird.

5.2. Neue Blockpermutationen mit Random Shift

[Politis & Romano \(1991\)](#) beschreiben in ihrem Technical Report eine Variante zur Verbesserung des moving bootstrap-Verfahrens. Denn beim ursprünglichen moving-bootstrap-Prinzip sind die ersten und letzten Elemente der in Blöcke zu zerlegenden Reihe (5.2) unterrepräsentiert, da sie ausschließlich an den Rändern der Blöcke auftauchen. Durch die Verbindung des Vektors mit sich selbst zu einem Kreis können an diesen Randstellen in gleicher Weise Blöcke, wie zum Beispiel $\{y_n, y_1, y_2, \dots, y_{l-1}\}$ oder $\{y_{n-l+3}, \dots, y_{n-1}, y_n, y_1, y_2\}$, gebildet und später in die Bootstrap-Stichproben aufgenommen werden: ein sogenanntes circular moving bootstrap mit insgesamt n Blöcken der Länge l .

Das in dieser Arbeit vorgeschlagene Verfahren der Blockpermutation mit einer zufälligen Verschiebung – einem Random Shift – verfährt ähnlich, wenn auch durch eine andere Intention begründet und in unterschiedlicher Weise realisiert.

Wie in Abbildung 5.3 dargestellt, wird in jedem Permutationsschritt vor Einteilung der Zeitreihe in Blöcke ein Random Shift, eine zufällige Verschiebung der Länge a ($a \in \mathbb{Z}$, $0 \leq a < n$) durchgeführt. Das bedeutet, a Elemente vom Anfang der Zeitreihe werden entfernt und in gleicher Reihenfolge an ihrem Ende angehängt. Dann erst wird die verschobene Zeitreihe in Blöcke eingeteilt und diese werden permutiert. Dabei wird a innerhalb des eben angegebenen Wertebereichs mittels einer Gleichverteilung simuliert. Im Prinzip entspricht dieser erste Schritt dem Vorgehen, bei dem die Zeitreihe wie in [Po-](#)

Blockweise Permutation mit zufälliger Verschiebung

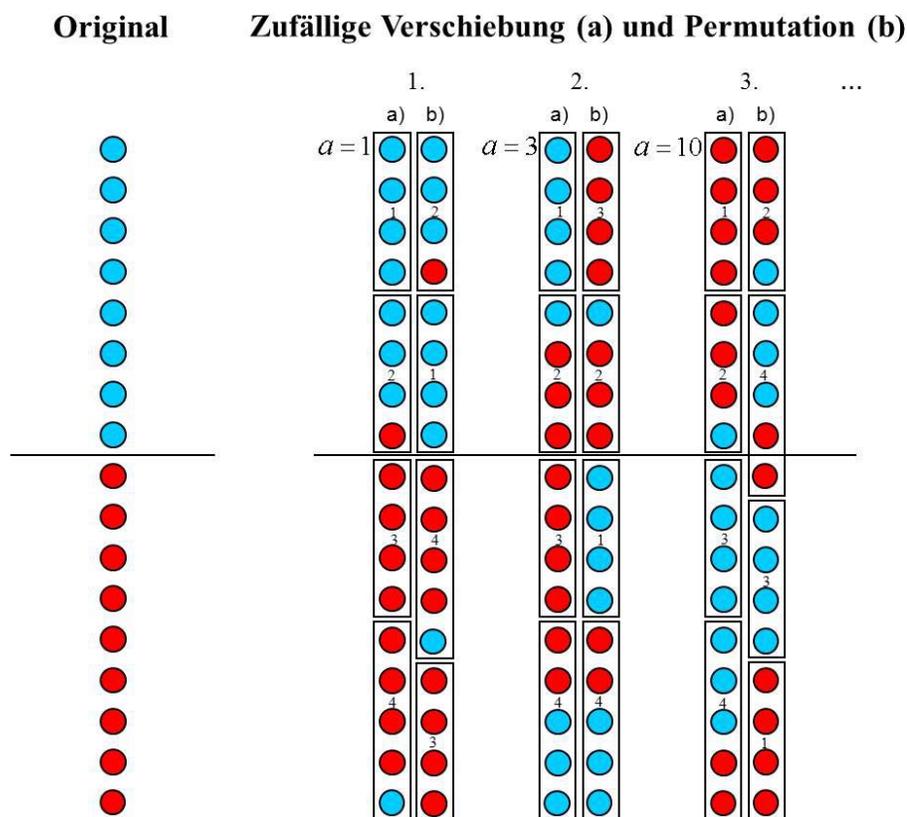


Abbildung 5.3.: Schematische Darstellung der blockweisen Permutation eines univariaten Zweistichprobenproblems mit Blocklänge $l = 4$, Blockanzahl $k = 4$ und zufälliger Verschiebung um den Wert a

litis & Romano (1991) zu einem Kreis zusammengefügt, dann um eine zufällige Anzahl a von Elementen gedreht und schließlich wieder geöffnet wird. Auch hier würden sich die zuvor a ersten Elemente der Zeitreihe in gleicher Reihenfolge am Ende des Vektors befinden.

Da dieser Random Shift in jedem einzelnen Permutationsschritt erfolgt und die Messreihe erst anschließend in k Blöcke aufgeteilt und permutiert wird, handelt es sich dabei jedesmal um unterschiedliche Blöcke. Dadurch kann die Anzahl voneinander ungleicher Permutationen erhöht werden, weil diese nun nicht mehr allein von der festen Anzahl k der Blöcke bestimmt wird.

Ist die Blocklänge l Teiler des Stichprobenumfangs, kann die Anzahl verschiedener Blockpermutationen um den Faktor l erhöht werden. Ist die Blocklänge kein Teiler des Stichprobenumfangs und damit der jeweils letzte Block länger als die restlichen, vergrößert sich die Anzahl der Permutationsmöglichkeiten sogar um den Faktor n .

Für eine Messreihe aus 15 Elementen und eine Blocklänge von $l = 5$ entstünden ohne

Random Shift genau drei voneinander verschiedene Blöcke und somit maximal sechs Kombinationsmöglichkeiten. Mit der zufälligen Verschiebung erhöht sich die Anzahl möglicher Kombinationen unterschiedlicher Blöcke bereits auf 30.

Durch den Random Shift ist beim Einsatz des Verfahrens für funktionelle Bilddaten zusätzlich gewährleistet, dass die Blockeinteilung nicht zufällig mit der Einteilung von Stimulations- und Ruhephasen übereinstimmt und dies über die gesamte Permutation gilt. Dadurch würden stets ausschließlich die gleichen Blöcke entsprechend einer Phase miteinander permutiert, was einen starken Güteverlust zur Folge hätte.

Eine erste Gegenüberstellung des Blockpermutationsprinzips ohne und mit zufälliger Verschiebung findet sich in [Adolf *et al.* \(2011\)](#).

Für die in dieser Arbeit durchgeführten Blockpermutationen wird stets ein Random Shift verwendet. Entsprechend den Ausführungen in Abschnitt [2.3.3](#) werden auch die Blockpermutationen mit zufälliger Verschiebung auf das transformierte Modell ([2.8](#)) und somit auf die die Nullhypothese betreffenden Teile der transformierten Designmatrix angewandt.

5.3. Auswahl geeigneter Blocklängen

Bei der Gegenüberstellung der Testverfahren anhand simulierter fMRT-Daten im nachfolgenden Kapitel wird als Adjustierungsmethode für korrelierte Stichprobenvektoren auch jeweils eine Blockpermutation mit zufälliger Verschiebung angegeben. Dabei dienen die in Abschnitt [2.3](#) beschriebenen Teststatistiken in den Permutationen als unterliegende Teststatistik. Für die Vielzahl möglicher Blocklängen ergibt sich damit auch eine Vielzahl von Blockpermutationsstatistiken, für die es wenig sinnvoll ist, sie alle stets den parametrischen Testverfahren gegenüberzustellen, da eine approximative Unabhängigkeit der Blöcke nicht bei allen Blocklängen erreicht werden kann. Daher soll in diesem Abschnitt für die verschiedenen Teststatistiken als Basis der Permutation ein Vergleich der Blocklängen bezüglich des Fehlers 1. Art erfolgen. Die minimale Blocklänge, für die das nominale Testniveau jeweils approximativ eingehalten werden kann, wird im nachfolgenden Kapitel den parametrischen Teststatistiken gegenübergestellt.

Es werden simulierte fMRT-Daten sowohl entsprechend eines Blockparadigmas wie auch in Form eines ereignisbezogenen Designs untersucht (vgl. Abschnitt [3.2.2](#)). Die Designs spiegeln sich in der jeweiligen Designmatrix wider, während die Daten (entsprechend 250 Scans über 1024 Voxel) unter der Nullhypothese mit räumlichen wie auch zeitlichen Abhängigkeiten simuliert werden. Weitere Details zur Simulation der Daten sind im Abschnitt [6.1](#) aufgeführt.

Es werden für jeden Test 2500 Wiederholungen mit je 299 zufälligen (blockweisen) Permutationen durchgeführt. Abweichungen innerhalb des Intervalls $[0, 0415; 0, 0585]$ können somit als rein zufällig betrachtet werden. Die Permutation erfolgt wie in Abschnitt [5.2](#) beschrieben innerhalb des interessierenden Teils der Designmatrix. Dieser besteht für beide Paradigmen aus der Differenz zweier Stimuli.

Im Folgenden werden die verschiedenen Blocklängen der Blockpermutation mit Random Shift bei Verwendung der in Abschnitt 2.3 vorgestellten Teststatistiken in Bezug auf die Einhaltung des nominalen Testniveaus miteinander verglichen. Vergleiche mit den parametrischen Verfahren sowie Untersuchungen zur Güte erfolgen dann wie bereits erwähnt im nächsten Kapitel anhand der in diesem Abschnitt ausgewählten Blocklänge.

In den folgenden Abbildungen wird der empirische Fehler 1. Art für die Blockpermutation mit Random Shift auf Basis der verschiedenen multivariaten Teststatistiken in Abhängigkeit vom zeitlichen AR(1)-Korrelationskoeffizienten ρ dargestellt. Auf der linken Seite der Abbildungen befinden sich jeweils die Ergebnisse für ein fMRT-Blockdesign, rechts für ein ereignisbezogenes Design. Die roten Balken verdeutlichen das Testverhalten bei klassischer elementweiser Permutation (Blocklänge = 1), also ohne Berücksichtigung der Korrelation der Stichprobenvektoren. Die übrigen Balken spiegeln Ergebnisse blockweiser Permutationen mit Random Shift wider. Dabei werden jeweils aufsteigend die folgenden Blocklängen betrachtet: Blocklänge fünf – dunkelblaue Balken; zehn – gelbe Balken; 15 – grün; 20 – lila; 40 – orange und Blocklänge 60 – schwarze Balken.

Selbst für die größte Blocklänge von 60 aufeinanderfolgenden Stichprobenvektoren ergeben sich mit vier Blöcken und dem Random Shift theoretisch noch 5976 Kombinationsmöglichkeiten. Für die hier simulierte Anzahl an Stichprobenvektoren ($n = 250$) existieren theoretisch auch bei nur drei Blöcken (ab Blocklänge = 63) noch 1494 Kombinationsmöglichkeiten. Da Permutationen innerhalb von Teilstichproben jedoch keine Veränderung der Teststatistik bewirken (vgl. Abschnitt 2.3.3), und dies vor allem beim Blockparadigma trotz der zufälligen Verschiebung nicht prinzipiell ausgeschlossen werden kann, könnte sich die Zahl verschiedener Kombinationsmöglichkeiten noch verringern. Deshalb werden für diesen Simulationsaufbau nur Blocklängen bis maximal 60 aufeinanderfolgende Stichprobenvektoren untersucht.

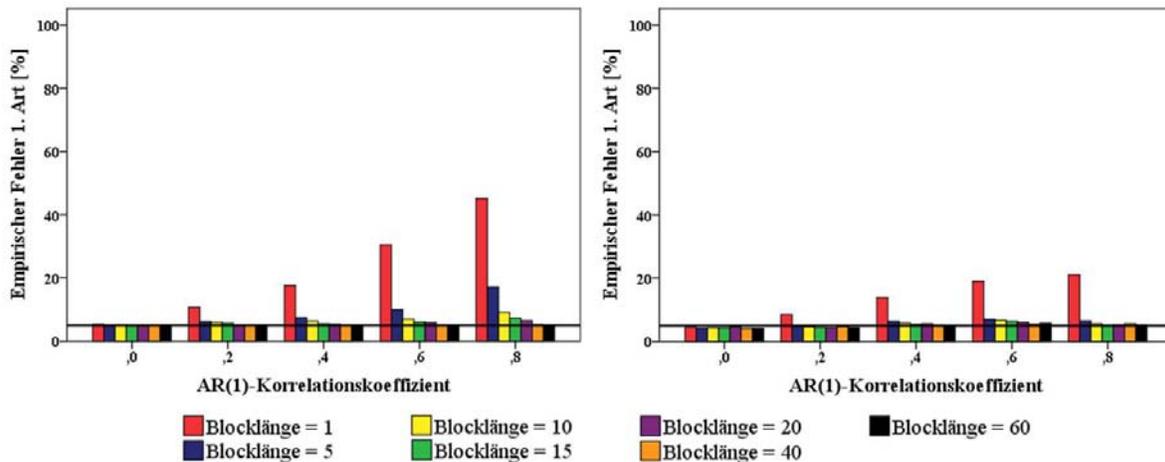


Abbildung 5.4.: Empirischer Fehler 1. Art der Blockpermutation mit Random Shift auf Basis der Standardisierten Summenteststatistik bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

Abbildung 5.4 zeigt die Simulationsergebnisse für Permutationen auf Grundlage der **Standardisierten Summenteststatistik**. Bei der klassischen Permutation (Blocklänge=1) ist erkennbar, dass der empirische Fehler 1. Art steigt, je höher die zeitliche Korrelation der Daten ist. Er erreicht für das Blockdesign (links) schließlich bei $\rho = 0,8$ etwa 45 Prozent. Beinahe jede zweite Nullhypothese wird bei dieser Korrelation der Stichprobenvektoren demnach fälschlicherweise abgelehnt.

Wird hingegen eine blockweise Permutation verwendet, verringert sich der empirische Fehler 1. Art mit wachsender Blocklänge. Bei einer zeitlichen Korrelation von $\rho = 0,2$ kann das nominale Testniveau bereits bei einer Blocklänge von zehn Stichprobenvektoren approximativ eingehalten werden. Bei einem mittleren Korrelationskoeffizienten von $\rho = 0,4$ ist dies bei einer Blocklänge von 20 Elementen möglich; für höhere Korrelationen bis zum maximal simulierten Wert $\rho = 0,8$ bei 40 Elementen je Block.

Beim ereignisbezogenen Design (rechts) ist die Verletzung des nominalen Testniveaus bei klassischer Permutation auf Basis der Standardisierten Summenteststatistik ebenso vorhanden, aber nicht so stark ausgeprägt wie beim Blockdesign. Der empirische Fehler 1. Art erreicht bei $\rho = 0,8$ einen Wert von etwa 21 Prozent. Entsprechend schneller kann das Testniveau hier mit wachsender Blocklänge approximativ eingehalten werden. Selbst bei hohen zeitlichen Korrelationen reichen dafür bei diesem funktionellen Paradigma bereits Blocklängen von 20 Stichprobenvektoren aus.

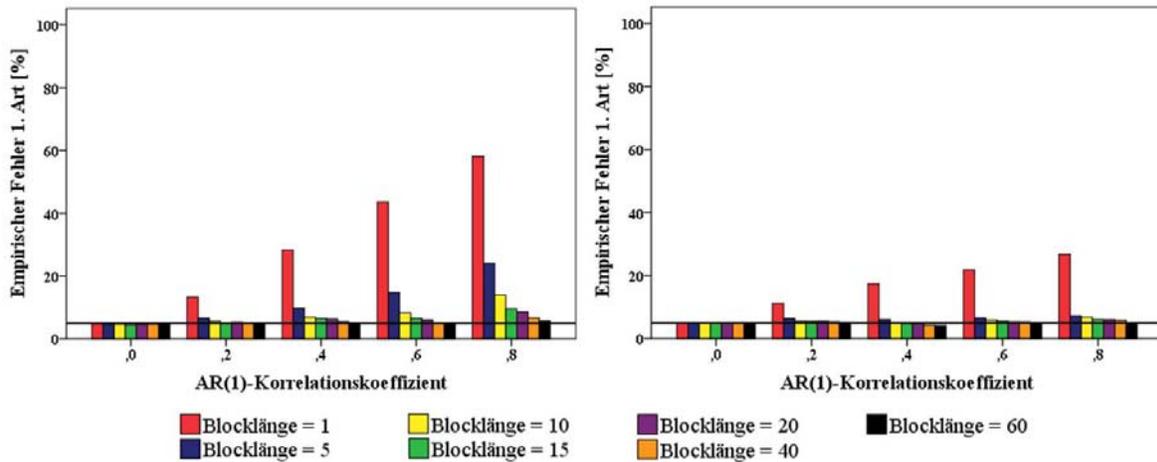


Abbildung 5.5.: Empirischer Fehler 1. Art der Blockpermutation mit Random Shift auf Basis der eindimensionalen skalenabhängigen Principal Component Teststatistik bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

Ähnlich verhält es sich, wenn die Blockpermutation mit Random Shift auf der Teststatistik des **eindimensionalen skalenabhängigen Principal Component Tests** basiert (Abbildung 5.5). Der Fehler 1. Art steigt für das Blockparadigma (links) bei klassischer Permutation auf Basis der skalenabhängigen PC-Teststatistik mit wachsendem Korrelationskoeffizienten auf knapp 60 Prozent an und liegt damit noch ein wenig

über den Ergebnissen bei Permutation auf Basis der Standardisierten Summenteststatistik. Auch hier nimmt die Verletzung des Testniveaus mit wachsender Blocklänge ab. So kann bei hohen zeitlichen Korrelationen ein empirischer Fehler in Höhe des nominalen Testniveaus durch Blocklängen von 40 beziehungsweise 60 aufeinanderfolgende Stichprobenvektoren erreicht werden.

Für das ereignisbezogene Design (rechts) wird bei klassischer Permutation auf Basis der eindimensionalen skalenabhängigen PC-Teststatistik bei $\rho = 0,8$ etwa jede dritte Nullhypothese fälschlicherweise abgelehnt. Permutationen mit Blöcken von mindestens 20 aufeinanderfolgenden Stichprobenvektoren erzielen jedoch bereits für alle untersuchten zeitlichen Korrelationsstärken ein approximatives Einhalten des Testniveaus.

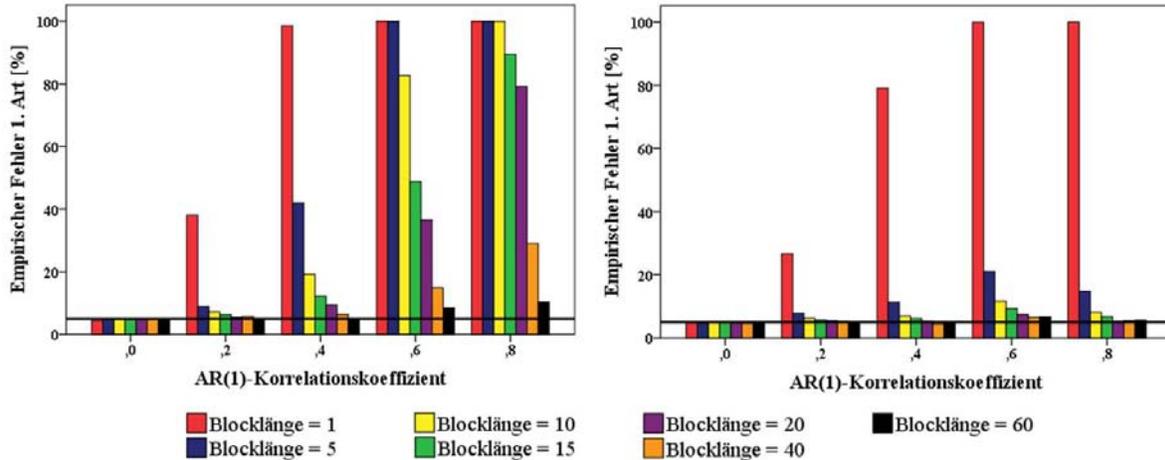


Abbildung 5.6.: Empirischer Fehler 1. Art der Blockpermutation mit Random Shift auf Basis der q -dimensionalen skalenabhängigen Principal Component Teststatistik bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

Etwas anders verhält es sich, wenn die dem Blockpermutationsverfahren zugrunde liegende PC-Teststatistik auf einem mehrdimensionalen Score basiert. In Abbildung 5.6 (links Blockdesign, rechts ereignisbezogenes Design) wird die Dimension q für den skalenabhängigen PC-Score aus den Daten anhand des inversen Jolliffe-Kriteriums (2.22) geschätzt. Bedingt durch die räumliche Block-Compound-Symmetry-Struktur der Daten (vgl. Abschnitt 6.1) wird für die Dimension im zeitlich unkorrelierten Fall der zu erwartende Wert $q = 16$ auch tatsächlich angenommen. Mit wachsendem ρ steigt die Schätzung jedoch auf bis zu 34 Hauptkomponenten an. Wie schon in Kapitel 3 dargestellt, bedingt die Dimensionsvergrößerung aufgrund der Korrelationen der Stichprobenvektoren eine zusätzliche Vergrößerung des empirischen Fehlers 1. Art.

Schon bei mittleren zeitlichen Korrelationen werden für das Blockdesign bei klassischer elementweiser Permutation alle wahren Nullhypothesen fälschlicherweise abgelehnt. Auch hier kann die Verletzung des Testniveaus zwar mit steigender Blocklänge durch die Blockpermutation mit zufälliger Verschiebung reduziert werden. Doch bereits

bei $\rho = 0,4$ sind Blöcke bestehend aus mindestens 60 aufeinanderfolgenden Stichprobenvektoren notwendig, um das Testniveau approximativ einzuhalten. Steigt die zeitliche Korrelation der Daten auf $\rho = 0,8$, reichen selbst diese Blocklängen nicht mehr aus. Für eine Blocklänge von 60 Stichprobenvektoren ergibt sich für simulierte fMRT-Daten eines Blockparadigmas noch ein empirischer Fehler 1. Art von etwa zehn Prozent.

Unterliegt die Simulation einem ereignisbezogenen Design, werden bei klassischer Permutation für hohe Korrelationen zwar ebenso ausschließlich falsch-positive Ergebnisse erzielt. Die blockweise Permutation mit Random Shift kann dies aber beim ereignisbezogenen Design mit wachsender Blocklänge beheben, so dass das Testniveau auch auf Basis **skalenabhängiger höherdimensionaler PC-Scores** mit Blocklängen ab 20 Stichprobenvektoren für alle untersuchten Korrelationskoeffizienten approximativ eingehalten werden kann.

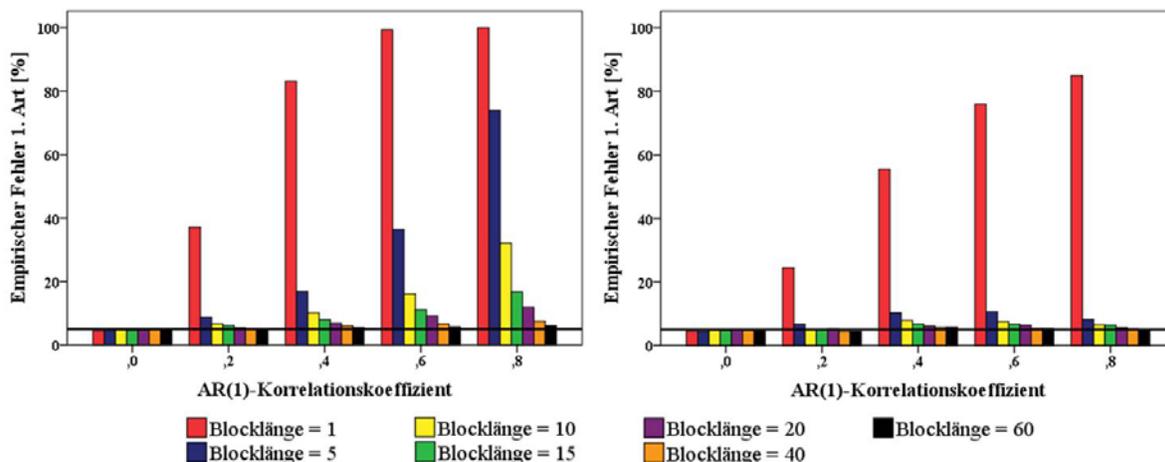


Abbildung 5.7.: Empirischer Fehler 1. Art der Blockpermutation mit Random Shift auf Basis der modifizierten q -dimensionalen skalenabhängigen Principal Component Teststatistik bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

Abbildung 5.7 zeigt, dass die Verletzung des nominalen Testniveaus bei klassischer Permutation nicht ganz so stark ausfällt, wenn die **modifizierte Form der q -dimensionalen PC-Teststatistik** genutzt wird. Mit steigender Blocklänge sinkt der empirische Fehler 1. Art dementsprechend schneller und kann bei einer Blocklänge von $l = 60$ beim ereignisbezogenen Design für alle untersuchten Korrelationsstärken, beim Blockdesign bis einschließlich $\rho = 0,6$ approximativ eingehalten werden. Bei $\rho = 0,8$ steigt der Fehler beim fMRT-Blockparadigma auf 6,1 Prozent an.

Bei Verwendung einer **skaleninvarianten Principal Component Teststatistik** als Basis der blockweisen Permutationen werden sowohl beim Blockdesign als auch beim ereignisbezogenen Design Ergebnisse erzielt, die nur geringfügig von denen mit skalenabhängiger Teststatistik abweichen. Die Simulationsergebnisse der (blockweisen) Permutationen auf Basis der skaleninvarianten PC-Teststatistiken (ein- und mehrdimensional

wie auch in modifizierter Form) werden daher hier nicht separat als Grafik angegeben, sondern in Anhang C tabellarisch dokumentiert.

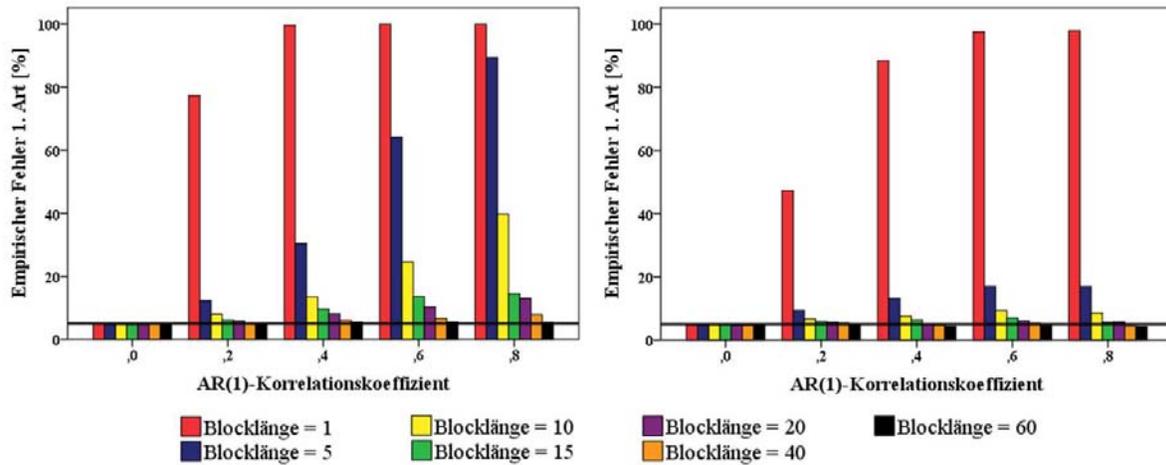


Abbildung 5.8.: Empirischer Fehler 1. Art der Blockpermutation mit Random Shift auf Basis eines Ähnlichkeitsmaßes in Form des Korrelationskoeffizienten nach Pearson bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

In Abbildung 5.8 ist der empirische Fehler 1. Art der Blockpermutation mit Random Shift auf Basis eines Ähnlichkeitsmaßes in Form des **Korrelationskoeffizienten nach Pearson** dargestellt.

Für beide fMRT-Designs steigt der empirische Fehler 1. Art mit wachsender zeitlicher Korrelation relativ schnell an. Für eine zeitliche Abhängigkeit von $\rho = 0,2$ werden bereits 77 beziehungsweise 47 Prozent der wahren Nullhypothesen abgelehnt. Ab $\rho = 0,6$ ergeben sich für beide fMRT-Paradigmen fast nur noch falsch-positive Resultate. Im Gegensatz zum Blockpermutationsverfahren auf Basis der mehrdimensionalen PC-Teststatistik (vgl. Abb. 5.6) kann diese Verletzung des nominalen Testniveaus hier auch beim Blockdesign (links) schon mit geringeren Blocklängen verringert werden. Während in Abbildung 5.6 für $\rho = 0,8$ bei einer Blocklänge von zehn Stichprobenvektoren noch alle Hypothesen fälschlicherweise abgelehnt werden, beträgt der empirische Fehler 1. Art für diese Blockpermutation auf Basis des Pearson'schen Korrelationskoeffizienten noch knapp 40 Prozent. Bei einer Blocklänge von 60 Elementen wird das Testniveau für alle simulierten Korrelationen approximativ eingehalten.

Beim ereignisbezogenen Design (rechts) hingegen kann die Verletzung des Fehlers 1. Art durch eine blockweise Permutation auf Basis des Korrelationskoeffizienten nach Pearson bereits durch geringere Blocklängen effektiv gesenkt werden. Bei mittleren zeitlichen Abhängigkeiten genügt eine Blockbildung mit 20 Elementen. Bei den untersuchten hohen Korrelationen der Stichprobenvektoren kann das Testniveau ab einer Blocklänge von 40 Elementen approximativ eingehalten werden.

Abbildung 5.9 enthält schließlich die Simulationsergebnisse der Blockpermutation mit

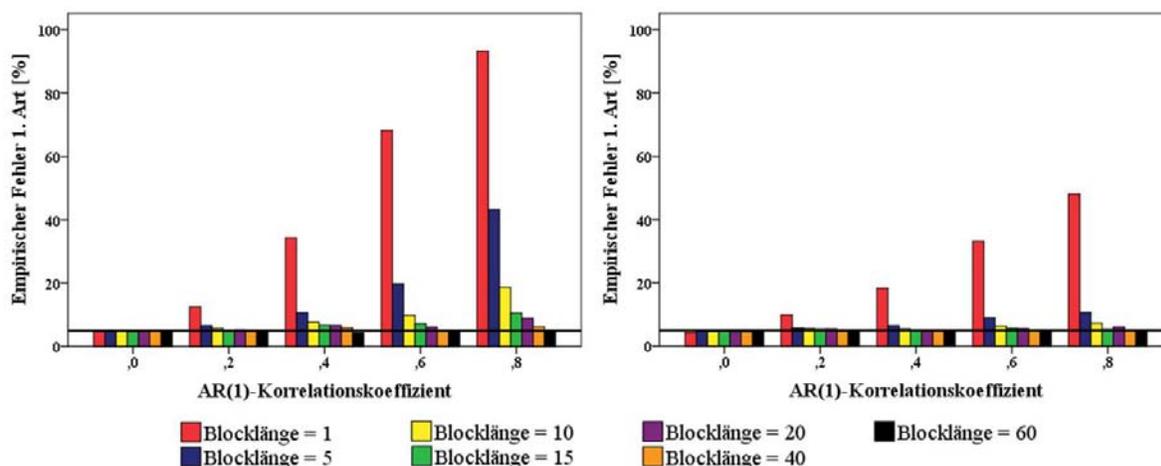


Abbildung 5.9.: Empirischer Fehler 1. Art der Blockpermutation mit Random Shift auf Basis eines Abstandsmaßes in Form des quadratischen euklidischen Abstandes bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

Random Shift auf Basis eines Abstandsmaßes in Form des **quadratischen euklidischen Abstandes** (QEA). Für das Blockparadigma (links) steigt der empirische Fehler 1. Art mit steigendem zeitlichen Korrelationskoeffizienten auf über 90 Prozent an. Durch eine blockweise Permutation mit zufälliger Verschiebung kann die Verletzung bei niedrigen Korrelationen bereits durch eine Blocklänge von zehn Stichprobenvektoren behoben werden. Ein approximatives Einhalten des nominalen Testniveaus für alle untersuchten zeitlichen Abhängigkeiten wird bei Blöcken mit mindestens 60 aufeinanderfolgenden Elementen erzielt.

Beim ereignisbezogenen Design (rechts) wird bei $\rho = 0,8$ durch klassische Permutation jede zweite Nullhypothese fälschlicherweise verworfen. Diese Verletzung des Testniveaus kann aber mit wachsender Blocklänge effektiv behoben werden. Selbst für hohe Korrelationen reichen Blocklängen von 40 Elementen aus, um das Testniveau approximativ einzuhalten.

Bei allen klassischen Permutationen auf Basis einer der untersuchten multivariaten Teststatistiken steigt der empirische Fehler 1. Art mit wachsender Korrelation der Stichprobenvektoren an. Bei einem fMRT-Blockdesign ist die Verletzung des nominalen Testniveaus durch autokorrelierte Stichprobenvektoren in allen Untersuchungen stärker als bei einem ereignisbezogenen Design.

In jedem Fall kann diese Verletzung des Fehlers 1. Art durch eine blockweise Permutation mit Random Shift mit steigender Blocklänge verringert werden. Für die Blockpermutation mit Random Shift auf Basis einer mehrdimensionalen PC-Teststatistik bei hohen zeitlichen Korrelationen erfolgt ein approximatives Einhalten des Testniveaus bis zu einer mittleren zeitlichen Korrelationen von $\rho = 0,4$. Abgesehen von dieser Testvariante kann das nominale Testniveau für alle Teststatistiken innerhalb der blockweisen Permu-

tation und für alle untersuchten AR(1)-Korrelationskoeffizienten bei einer Blocklänge von 60 Stichprobenvektoren approximativ eingehalten werden. Durch die zufällige Verschiebung ist dabei gewährleistet, dass die Verteilung der jeweiligen Teststatistik durch Permutationen auch bei nur vier Blöcken detailliert genug beschrieben werden kann, um eine Aussage bezüglich des Fehlers 1. Art zu treffen.

In den umfassenden Gegenüberstellungen der Adjustierungsmethoden für korrelierte Stichprobenvektoren im folgenden Kapitel wird daher für die blockweise Permutation mit Random Shift jeweils die Blocklänge 60 genutzt.

6. Gegenüberstellung der Verfahren anhand simulierter Daten

Um die in den vorherigen Kapiteln beschriebenen statistischen Verfahren auf funktionelle Magnetresonanztomographie-Daten anwenden zu können, muss zunächst gewährleistet sein, dass der Fehler 1. Art trotz der Korrelation der Stichprobenvektoren dem nominalen Testniveau entspricht. Unter dieser Voraussetzung ist anschließend von Interesse, welches Verfahren sich dabei als gütestärkstes erweist.

Hierzu werden zuerst alle multivariaten Teststatistiken für simulierte fMRT-Daten anhand der verschiedenen Adjustierungsverfahren für korrelierte Stichproben auf approximative Einhaltung des nominalen Testniveaus untersucht. Unter Einhaltung des Niveaus werden die Teststatistiken danach bezüglich ihrer Güte, des Anteils der richtig verworfenen Nullhypothesen, miteinander verglichen. Zusätzlich erfolgt der Vergleich der multivariaten Tests mit dem in der funktionellen Bildgebung etablierten univariaten Verfahren bei entsprechender multipler Adjustierung: einem t -Test unter Anwendung des Pre-Whitenings, um Vor- und Nachteile der multivariaten Verfahren herausstellen zu können.

6.1. Design der Simulationsstudie

Es werden beide Designvarianten der funktionellen Bildgebung, das Block- und das ereignisbezogene Design (vgl. Abschnitt 3.2.2), simuliert, da sich bei Experimenten für funktionelle Bilddaten diese beiden Hauptstrategien des Versuchsaufbaus bezüglich der Aktivierungserkennung und damit auch bezüglich ihrer Anforderungen an die Sensitivität von Testverfahren voneinander unterscheiden könnten. Die simulierten Designs sind realen Studien nachempfunden. Es werden 250 Messungen über 1024 Voxel simuliert und der untersuchte Kontrast enthält jeweils die Differenz zweier Stimuli.

Die Daten des Blockdesigns entsprechen einer funktionellen Messung mit zwei Stimuli (A und B) über insgesamt 250 Scans, die in Zwei-Sekunden-Abständen gemacht werden. Stimulus- und Ruhephasen dauern jeweils 20 Sekunden an und werden achtmal wiederholt. Dabei findet die Zuweisung des Stimulus A oder B zufällig statt.

Das ereignisbezogene Design besteht ebenso aus 250 Scans, gemessen in Zwei-Sekunden-Abständen, jedoch mit drei sich zufällig abwechselnden Stimuli (A, B und C), die jeweils eine Sekunde anhalten. Die Stimuli treten insgesamt 80 mal auf und sind durch Ruhe-

phasen von fünf bis sieben Sekunden zufälliger Länge voneinander getrennt.

Für beide Designs wird die Designmatrix erstellt, indem die Stimulusvektoren mit der in Abschnitt 3.2.2 beschriebenen hämodynamischen Antwortfunktion (3.2) gefaltet werden. Anschließend werden die Zeitpunkte entsprechend des Zwei-Sekunden-Abstandes einer üblichen fMRT-Messung ausgelesen, so dass die Matrix im Ergebnis mit den Scans synchronisiert ist und wie die Beobachtungsmatrix 250 Zeilen umfasst. Durch die Faltung können die Einträge aufgrund des post-stimulus undershoots der HRF und der Überlagerung der Stimuli die ursprünglichen Min- und Maximumwerte unter- beziehungsweise überschreiten. Zusätzlich geht jeweils ein Gesamtmittel in die Untersuchungen ein, indem ein Vektor bestehend aus Einsen in die Designmatrizen eingefügt wird.

Abbildung 6.1 zeigt die Darstellung der simulierten Regressoren für das Blockdesign (links) und für das ereignisbezogene Design (rechts). Die Grauwertskala ist jeweils durch den minimalen (schwarz) und maximalen (weiß) Wert in der Matrix festgelegt. Der Grauwert der ersten Spalte charakterisiert das Gesamtmittel und somit jeweils den Wert 1. Da sich der Wertebereich der Regressoreneinträge aus der Faltung der Onsetvektoren mit der Designmatrix ergibt und die Abfolge und damit auch die Überlagerung der Stimuli je Design verschieden sind, entstehen für beide Bilder unterschiedliche Grauwertskalen. Die Designmatrix für das Blockdesign besteht aus der Spalte für das Gesamtmittel sowie zwei Spalten entsprechend den Stimuli A und B. Beim ereignisbezogenen Design umfasst die Matrix eine Spalte für das Gesamtmittel und drei weitere Spalten für die Stimuli A, B und C. Diese Designmatrizen fließen dann in die jeweiligen Modelle und Schätzungen ein.

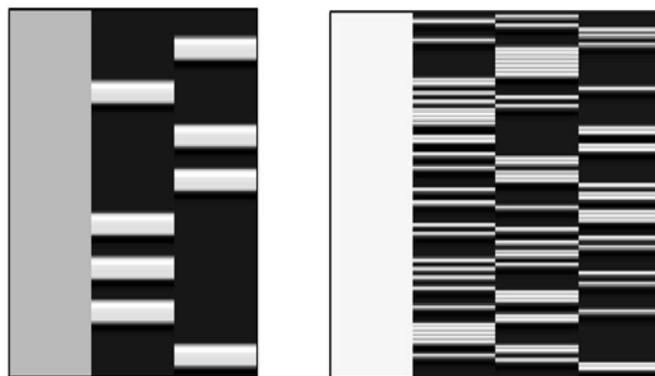


Abbildung 6.1.: Darstellung der Regressoren der Designmatrix für das Blockdesign (links; Gesamtmittel und Stimuli A und B) und das ereignisbezogene Design (rechts; Gesamtmittel und Stimuli A, B und C)

Die Simulation der Messungen unter der Nullhypothese erfolgt in zwei Schritten. Zunächst werden standardnormalverteilte Beobachtungen \mathbf{Y}_T der Dimension $n \times p$ ($n = 250$, $p = 1024$) erzeugt: $\mathbf{Y}_T \sim N(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{I}_p)$. Die Einarbeitung zeitlicher und räumlicher Kor-

relation erfolgt unter Ausnutzung der Eigenschaft der matrixvariaten Normalverteilung, derzufolge sich eine Multiplikation an die Daten von links ausschließlich auf die Kovarianzstruktur der Stichprobenvektoren, eine Multiplikation von rechts sich hingegen auf die Kovarianzstruktur der Variablen auswirkt (vgl. Satz 2.2). Somit entsteht die Matrix der korrelierten Daten \mathbf{Y} aus

$$\mathbf{Y} = \mathbf{P}^{\frac{1}{2}} \mathbf{Y}_I \boldsymbol{\Sigma}^{\frac{1}{2}}, \quad (6.1)$$

wobei \mathbf{P} der zeitlichen und $\boldsymbol{\Sigma}$ der räumlichen Abhängigkeitsstruktur entspricht und $\mathbf{P}^{\frac{1}{2}}$ beziehungsweise $\boldsymbol{\Sigma}^{\frac{1}{2}}$ die symmetrische Wurzel der jeweiligen Matrix verkörpert. Diese kann beispielsweise über die Eigenwertzerlegung der Matrix entsprechend Definition 2.7 gebildet werden. Damit ergibt sich für die Verteilung von \mathbf{Y}

$$\mathbf{Y} \sim N(\mathbf{P}^{\frac{1}{2}} \mathbf{0} \boldsymbol{\Sigma}^{\frac{1}{2}}, \mathbf{P}^{\frac{1}{2}} \mathbf{I}_n \mathbf{P}^{\frac{1}{2}'} \otimes \boldsymbol{\Sigma}^{\frac{1}{2}'} \mathbf{I}_p \boldsymbol{\Sigma}^{\frac{1}{2}})$$

beziehungsweise vereinfacht

$$\mathbf{Y} \sim N(\mathbf{0}, \mathbf{P} \otimes \boldsymbol{\Sigma}).$$

Dabei wird \mathbf{P} als Korrelationsmatrix eines autoregressiven Prozesses erster Ordnung entsprechend (3.7) für unterschiedliche Korrelationskoeffizienten $\rho \in \{0, 0, 0, 2; 0, 4; 0, 6; 0, 8\}$ simuliert. $\boldsymbol{\Sigma}$ wird als Block-Compound-Symmetry-Struktur nach (3.9) mit insgesamt 16 Blöcken zu je 64 Variablen gebildet. Innerhalb dieser Blöcke wird als Korrelationsstärke der Wert $\rho_r = 0,5$ gewählt.

Die Daten \mathbf{Y} zur Untersuchung des Fehlers 1. Art sind damit unter der globalen Nullhypothese simuliert, dass keine der potentiellen Einflussvariablen einen signifikanten Einfluss auf die Daten hat. Bei den anschließenden Gütesimulationen werden jedoch eingeschränktere Nullhypothesen untersucht, genauer gesagt die Differenz zweier Einflussvariablen geprüft. Das Ergebnis beim Test einer solchen paarweisen Nullhypothese wird aber nicht durch den konkreten Parameterwert der zu vergleichenden Stimuli beziehungsweise durch den Wert der nicht in der Nullhypothese betroffenen weiteren Stimuli verändert, solange die untersuchte Differenz der beiden Paarparameter tatsächlich Null bleibt. Eine Vereinfachung auf die globale Nullhypothese stellt somit keine Einschränkung dar und kann exemplarisch für die Untersuchung der Teststatistiken bezüglich des Fehlers 1. Art verwendet werden.

Für die Simulationen unter der Alternativhypothese wird auf die Datenmatrix \mathbf{Y} nach (6.1) eine Matrix \mathbf{M} addiert. Diese wird unter Nutzung der Spalten \mathbf{x}_i der jeweiligen Designmatrix gebildet ($i = 1, 2, 3$ beim Blockdesign: Gesamtmittel, Stimuli A und B; $i = 1, \dots, 4$ beim ereignisbezogenen Design: Gesamtmittel, Stimuli A, B und C) und ergibt sich somit für das Blockdesign aus

$$\mathbf{M} = \left(\mathbf{X} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \mathbf{1}'_{p_1} \quad \mathbf{0}_{n \times (p-p_1)} \right)$$

und für das ereignisbezogene Design aus

$$\mathbf{M} = \left(\mathbf{X} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} \mathbf{1}'_{p_1} \quad \mathbf{0}_{n \times (p-p_1)} \right).$$

\mathbf{M} entspricht daher jeweils einer $n \times p$ -dimensionalen Matrix, die in den ersten p_1 Spalten die gewichtete Summe der Regressoren und in den restlichen Spalten Nullvektoren enthält, wobei $p_1 < p$ zuvor festgelegt wird. Bei Addition dieser Matrix zu (6.1) beeinflussen somit die Regressoren die ersten p_1 Spalten der simulierten Daten.

Um eine Vergleichbarkeit der einzelnen Teststatistiken, Adjustierungsverfahren und zeitlichen Korrelationsstärken zu gewährleisten und die ersten p_1 Spalten der Beobachtungsmatrix in allen Simulationen in gleicher Weise unter der Alternativhypothese, nämlich der Differenz zweier Stimuli, zu stellen, werden in allen Simulationen durchgängen die folgenden Parametereinstellungen genutzt. Für das Blockdesign gilt

$$\beta_1 = 0, \quad \beta_2 = 0,14 \quad \text{und} \quad \beta_3 = 0,01$$

und für das ereignisbezogene Design gilt

$$\beta_1 = 0, \quad \beta_2 = 0,85, \quad \beta_3 = 0 \quad \text{und} \quad \beta_4 = 0,25.$$

Diese Effektstärken sind so gewählt, dass die gütestärksten multivariaten Statistiken für diese Einstellungen Gütewerte um 80 Prozent erreichen. Da die Güte jedoch für die verschiedenen multivariaten Teststatistiken, Adjustierungsverfahren und zeitlichen Korrelationsstärken sehr unterschiedlich hoch ausfällt, erfolgen einzelne Vergleiche auf niedrigerem Niveau.

Sowohl für die Untersuchungen zur Einhaltung des nominalen Testniveaus wie auch für den Gütevergleich der Verfahren werden jeweils 2500 Wiederholungen simuliert und darin gegebenenfalls je 299 Permutationen beziehungsweise Rotationen durchgeführt.

Im Fokus der Güteuntersuchungen steht die Situation, dass zwei der 16 räumlich voneinander unabhängigen Variablenblöcke unter der Alternative simuliert sind, also $p_1 = 128$. Weiterhin werden für eine mittlere zeitliche Korrelation von $\rho = 0,4$ Kontrollszenarien simuliert, um Vor- und Nachteile der Teststatistiken zu überprüfen.

Dabei wird zum einen eine Güteuntersuchung durchgeführt, der dieselben Parametereinstellungen zur Generierung der Daten unter der Alternative zugrunde liegen – mit dem Unterschied, dass hier nur $p_1 = 64$ Variablen, also ein einziger Variablenblock, unter der Alternativhypothese simuliert werden. Diese 64 Variablen sind somit über eine Compound-Symmetry-Struktur einheitlich miteinander korreliert und die Daten entsprechen fMRT-Aktivierungen in einem einzelnen funktionellen Hirnareal.

Zum anderen wird eine Simulation mit denselben Parametereinstellungen wie für die im Fokus stehende Alternative durchgeführt, bei der sich jedoch die räumliche Korrelation der 1024 Variablen in jedem Simulationsschritt vollkommen zufällig ergibt. Im Detail

bedeutet dies, dass die bei der Simulation der Daten nach (6.1) benutzte Korrelationsmatrix $\Sigma = (\sigma_{ij})$, $i, j = 1, \dots, p$, mit $\sigma_{ij} \in (-1, 1)$, $\sigma_{ii} = 1$ simuliert wird. Dazu wird zunächst eine iid standardnormalverteilte Matrix \mathbf{A} der Dimension $p \times p$ erzeugt, aus der dann über $\Sigma^{\frac{1}{2}} = \mathbf{A}(\text{diag}(\mathbf{A}'\mathbf{A}))^{-\frac{1}{2}}$ direkt die symmetrische Wurzel der Korrelationsmatrix gebildet werden kann. Im Kontext funktioneller Bilddaten ist diese unstrukturierte räumliche Korrelation eher sehr unwahrscheinlich. Doch damit soll in den anschließenden Güteuntersuchungen im Vergleich zur Annahme funktioneller Gebiete ein eventueller Vorteil des Hauptkomponentenverfahrens bei dieser konkreten Annahme eingeschätzt und somit verdeutlicht werden, wie sehr das Verhalten der adjustierten Testverfahren von der angenommenen räumlichen Korrelationsstruktur der Daten beeinflusst ist.

Schließlich wird ebenso untersucht, wie gut die Verfahren für simulierte fMRT-Daten bei einer extrem verkürzten Zeitreihe ($n = 50$) aber ansonsten unveränderten Parametereinstellungen funktionieren. Für das ereignisbezogene Design ergibt sich diese verkürzte Beobachtungsreihe aus den ersten 50 Zeitpunkten des in Abschnitt 6.1 beschriebenen Simulationsszenarios (also den ersten 50 Zeilen der Beobachtungs- wie auch Designmatrix). Beim Blockdesign werden hierfür die Zeitpunkte 16 bis 65 ausgewählt. Da sowohl Stimulus- als auch Ruhephasen in diesem Design 20 Sekunden andauern, ist somit gewährleistet, dass beide Stimuli in der ausgewählten Messreihe enthalten sind.

6.2. Empirischer Fehler 1. Art

Die verschiedenen Adjustierungsverfahren für die zeitliche Korrelation (Satterthwaite-Approximation, Pre-Whitening sowie die blockweise Permutation mit Random Shift, vgl. Kapitel 4 und 5) werden für die in Kapitel 2 vorgestellten multivariaten Teststatistiken zunächst auf das approximative Einhalten des nominalen Testniveaus untersucht. Parallel zur multivariaten Datenanalyse erfolgt die Durchführung der univariaten Tests anhand einer t -Teststatistik entsprechend (2.12) bei vorherigem Pre-Whitening.

Dabei werden alle Tests unter der Nullhypothese mit einer Irrtumswahrscheinlichkeit von $\alpha = 0,05$ durchgeführt. Aufgrund der Simulationsanzahl von 2500 Wiederholungen liegen bei exaktem Einhalten des Testniveaus 95 Prozent der Ergebnisse innerhalb des Intervalls $[0,0415; 0,0585]$.

6.2.1. Ergebnisse der simulierten fMRT-Daten

Die Ergebnisse der **univariaten Analysen** für unter der Nullhypothese simulierte fMRT-Daten sind in Tabelle 6.1 in Abhängigkeit von der zeitlichen Korrelationsstärke ρ aufgeführt. Der empirische Fehler 1. Art spiegelt den prozentualen Anteil der t -Tests wider, in denen unter der Nullhypothese in mindestens einer der 1024 Variablen eine signifikante Aktivierung nachgewiesen wird. Dabei wird das nach Bonferroni adjustierte multiple Testniveau von $\frac{0,05}{1024}$ genutzt. Dies gewährleistet, dass die sogenannte familywise error rate (FWER), die Wahrscheinlichkeit bei Durchführung mehrerer Test mindestens einen Fehler 1. Art zu begehen, maximal der Irrtumswahrscheinlichkeit $\alpha = 0,05$ entspricht.

Design	ρ				
	0,0	0,2	0,4	0,6	0,8
Blockdesign	0,041	0,046	0,050	0,069	0,062
ereignisbezogenes Design	0,039	0,047	0,040	0,043	0,048

Tabelle 6.1.: Empirischer simultaner Fehler 1. Art der t -Teststatistiken für simulierte fMRT-Daten bei Durchführung von Pre-Whitening zum Nachweis mindestens einer signifikanten Aktivierung ($n = 250$, $p = 1024$, Bonferroni-korrigiert, 2500 Wiederholungen)

Die Ergebnisse verdeutlichen, dass das nominale Testniveau beim ereignisbezogenen Design für sämtliche untersuchten zeitlichen Korrelationen approximativ nicht überschritten wird. Das Verfahren ist aufgrund der Bonferroni-Adjustierung sogar eher etwas konservativ.

Beim Blockdesign wird dieses Verhalten bis zu mittleren zeitlichen Korrelationsstärken ebenso beobachtet. Für $\rho \geq 0,6$ sind die Ergebnisse mit bis zu 6,9 Prozent für den empirischen Fehler 1. Art trotz Pre-Whitening noch leicht liberal, was bei den späteren Güteuntersuchungen berücksichtigt wird.

Dieses leicht liberale Verhalten bei hohen zeitlichen Korrelationsstärken und zugrunde liegendem fMRT-Blockdesign ist weiterhin auch für einige der multivariaten Teststatistiken zu beobachten. Eine ungenaue Schätzung des zeitlichen Korrelationskoeffizienten beim Pre-Whitening kann jedoch nicht der alleinige Grund dieser Überschätzung sein, da die Schätzungen für die beiden fMRT-Designs vergleichbar sind: beispielsweise wird für $\rho = 0,8$ beim Blockdesign $\hat{\rho} = 0,7890$ ($\text{RMSE}_{\hat{\rho}} = 0,0122$) und beim ereignisbezogenen Design $\hat{\rho} = 0,7869$ ($\text{RMSE}_{\hat{\rho}} = 0,0141$) erreicht.

In den folgenden Abbildungen wird der empirische Fehler 1. Art für unter der Nullhypothese simulierte funktionelle MRT-Daten bei Verwendung je einer der untersuchten multivariaten Teststatistiken in Abhängigkeit vom zeitlichen AR(1)-Korrelationskoeffizienten ρ dargestellt. Dies erfolgt für die originale Teststatistik im Vergleich zu den unterschiedlichen Adjustierungsverfahren für zeitliche Korrelation.

Auf der linken Seite der Abbildungen befinden sich jeweils die Ergebnisse für das fMRT-Blockdesign, rechts für das ereignisbezogene Design. Die roten Balken verdeutlichen das Testverhalten bei klassischer Verwendung der jeweiligen Teststatistik, also ohne Berücksichtigung der Korrelation der Stichprobenvektoren. Die übrigen Balken spiegeln Ergebnisse der Adjustierungsstrategien wider (Satterthwaite-Approximation – dunkelblau; Pre-Whitening – gelb; Blockpermutation mit Random Shift – grün). Beim Blockpermutationsverfahren wird stets die Blocklänge $l = 60$ verwendet (vgl. Abschnitt 5.3). Die Farbgebung wird nachfolgend in allen Diagrammen beibehalten.

In Abbildung 6.2 ist der empirische Fehler 1. Art des **Standardisierten Summentests** dargestellt. Ohne Berücksichtigung der zeitlichen Abhängigkeitsstruktur steigt der Fehler für die untersuchten Korrelationsstärken auf über 45 Prozent beim Blockdesign (links)

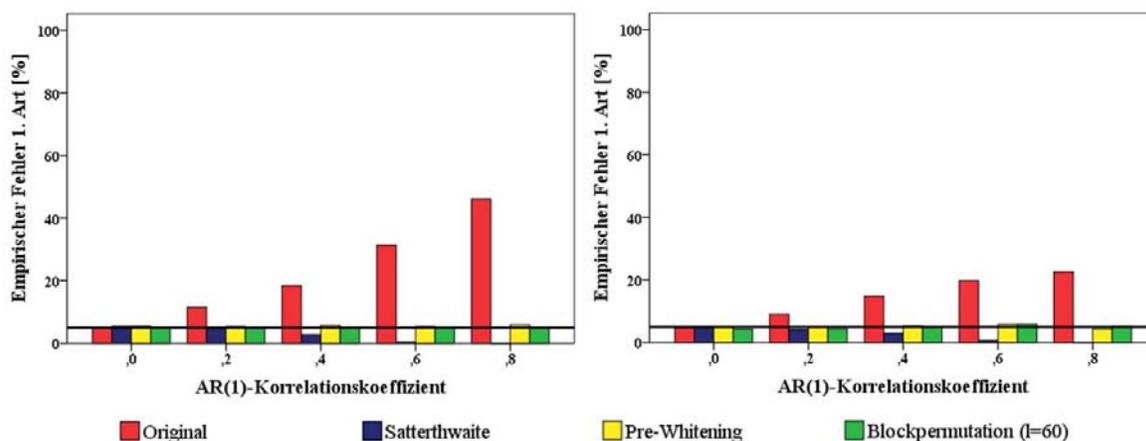


Abbildung 6.2.: Empirischer Fehler 1. Art des Standardisierten Summentests bei simulierten fMRT-Daten für verschiedene Adjustierungsverfahren; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

und knapp 25 Prozent beim ereignisbezogenen Design (rechts) an. Demnach würde fast jede zweite beziehungsweise jede vierte wahre Nullhypothese fälschlicherweise abgelehnt werden. Bei Verwendung der verschiedenen Adjustierungsmethoden für die autokorrelierten Stichprobenvektoren kann das nominale Testniveau aber für beide fMRT-Designs und alle untersuchten AR(1)-Korrelationskoeffizienten approximativ eingehalten werden. Allerdings werden die Ergebnisse der Satterthwaite-Approximation mit zunehmendem ρ konservativ und liegen für höhere zeitliche Korrelationen unter einem Prozent.

Die Simulationsergebnisse für den empirischer Fehler 1. Art des **eindimensionalen**

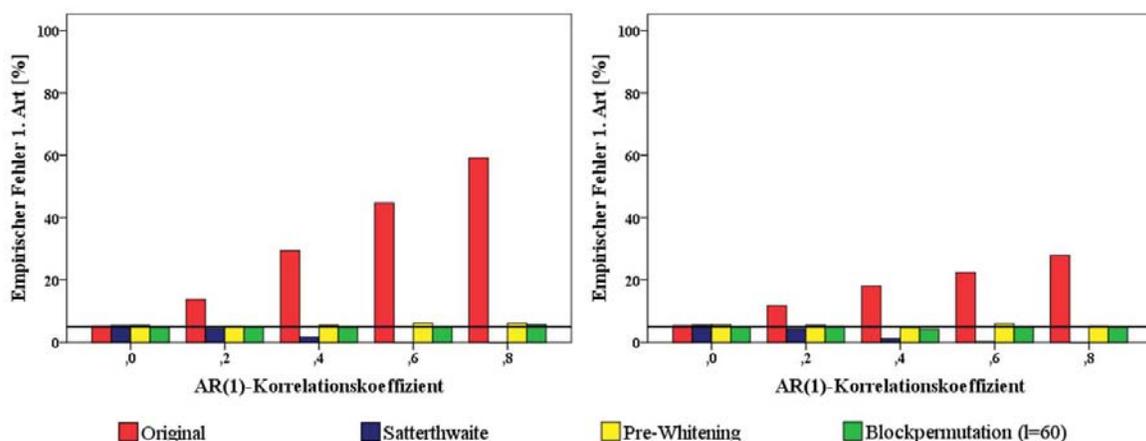


Abbildung 6.3.: Empirischer Fehler 1. Art des eindimensionalen skalenabhängigen Principal Component Tests bei simulierten fMRT-Daten für verschiedene Adjustierungsverfahren; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

skalenabhängigen Principal Component Tests sind in Abbildung 6.3 dargestellt. Sie gleichen denen der eben beschriebenen Standardisierten Summenteststatistik. Ohne Beachtung der Autokorrelation der Stichprobenvektoren werden hier gut 50 Prozent (Blockdesign; links) beziehungsweise ein Viertel (ereignisbezogenes Design; rechts) falsch-positive Ergebnisse erzielt. Anhand der parametrischen Adjustierungsmethoden wie auch der Blockpermutation mit Random Shift kann das nominale Testniveau größtenteils approximativ eingehalten werden. Lediglich beim Blockdesign ist das Pre-Whitening für hohe zeitliche Korrelationskoeffizienten mit maximal 6,4 Prozent leicht liberal, wohingegen die Satterthwaite-Approximation erneut für beide Designs mit steigendem ρ sehr konservativ wird.

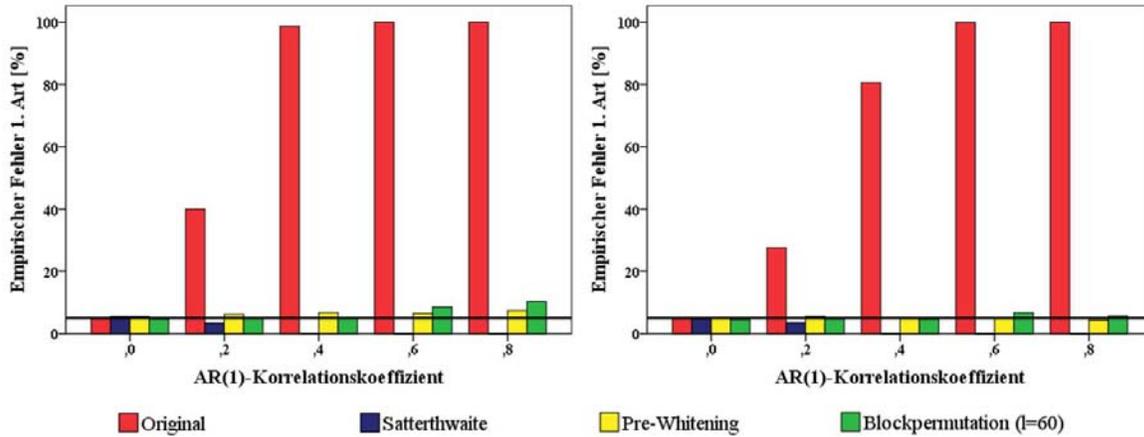


Abbildung 6.4.: Empirischer Fehler 1. Art des q -dimensionalen skalenabhängigen Principal Component Tests bei simulierten fMRT-Daten für verschiedene Adjustierungsverfahren; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

Die Verletzung des Testniveaus beim unkorrigierten **q -dimensionalen skalenabhängigen Principal Component Test** (Abbildung 6.4) ist wesentlich höher als beim eindimensionalen Pendant. Die Dimension der PC-Scores entwickelt sich dabei in Abhängigkeit von der zeitlichen Korrelation bei beiden fMRT-Designs von $q = 16$ entsprechend der simulierten räumlichen Gebiete (bei $\rho \leq 0,2$) auf im Mittel $q = 20$ bei $\rho = 0,4$ und schließlich $q = 34$ beziehungsweise $q = 33$ für $\rho = 0,6$ und $\rho = 0,8$. Die verschiedenen Adjustierungsmethoden bewirken zwar auch hier eine deutliche Senkung der falsch-positiven Resultate, was beim ereignisbezogenen Design (rechts) für ein approximatives Einhalten des Testniveaus sorgt. Dabei fällt die Satterthwaite-Approximation wiederum sehr konservativ aus. Bei einem unterliegenden fMRT-Blockparadigma kann das Testniveau jedoch für mittlere und hohe Korrelationskoeffizienten bei Verwendung von Pre-Whitening oder Blockpermutation nicht eingehalten werden. Der empirische Fehler 1. Art steigt für die untersuchten Korrelationsstärken beim Pre-Whitening auf 7,44 Prozent, bei Verwendung der Blockpermutation mit Random Shift auf 10,3 Prozent an. Beim Pre-Whitening ist der Score für alle untersuchten zeitlichen Korrelationen

wieder 16-dimensional.

Die **skaleninvariante PC-Teststatistik** liefert, wie schon bei den Simulationen zur blockweisen Permutation in Kapitel 5, auch bei diesen Simulationsuntersuchungen nahezu gleiche Ergebnisse wie das entsprechende skalenabhängige Verfahren – für beide fMRT-Designs im Original wie auch für alle Adjustierungsmethoden. Analog zu den Ergebnissen in Abbildung 6.3 wird das Testniveau bei der eindimensionalen Teststatistik durch die Adjustierungsverfahren approximativ eingehalten. Für die q -dimensionale skaleninvariante PC-Teststatistik kann das nominale Testniveau analog zu Abbildung 6.4 beim ereignisbezogenen Design vollkommen beim Blockdesign bis zu mittleren Korrelationen durch Pre-Whitening oder Blockpermutation approximativ eingehalten werden und steigt dort anschließend auf 7,4 beziehungsweise zehn Prozent an, während die Satterthwaite-Approximation jeweils extrem konservativ wird. Die Ergebnisse des ein- wie auch q -dimensionalen skaleninvarianten Tests sind tabellarisch in Anhang C dokumentiert.

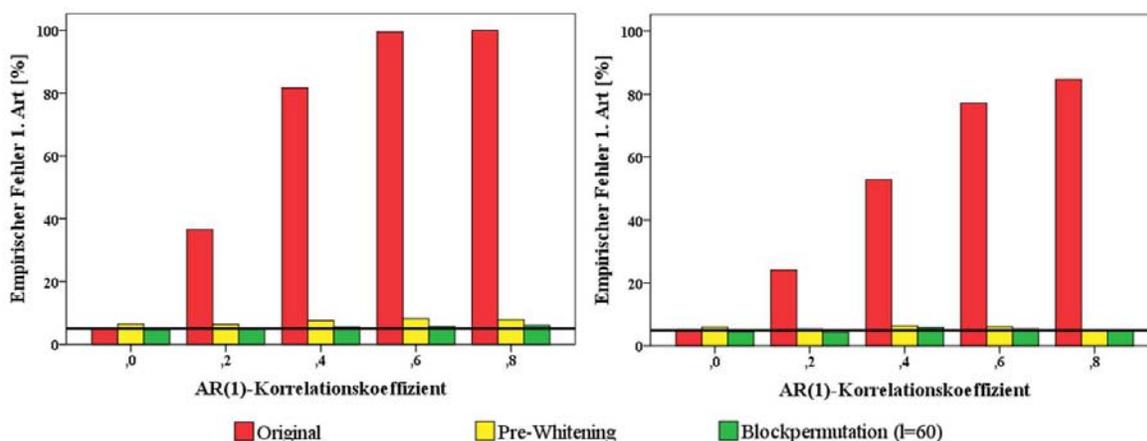


Abbildung 6.5.: Empirischer Fehler 1. Art des modifizierten q -dimensionalen skalenabhängigen Principal Component Tests bei simulierten fMRT-Daten für verschiedene Adjustierungsverfahren; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 299 Rotationen bzw. Permutationen, 2500 Wiederholungen)

Ähnlich wie die q -dimensionalen PC-Teststatistiken verhält sich auch der **modifizierte Principal Component Test**, bei dem die eigentlich q -dimensionalen Scores in eine eindimensionale Teststatistik umgerechnet werden. Wird die Verteilung der Teststatistik anhand eines Rotationsverfahrens generiert, kommt als Adjustierungsmethode für autokorrelierte Stichprobenvektoren das Pre-Whitening zum Einsatz. Bei Verwendung einer Permutationsverteilung kann die blockweise Permutation mit Random Shift realisiert werden. Der entsprechende empirische Fehler 1. Art für den Originaltest wie auch für die Adjustierungen sind in Abbildung 6.5 dargestellt.

Beim ereignisbezogenen Design (rechts) wird durch das Pre-Whitening ein approximati-

ves Einhalten des nominalen Testniveaus gewährleistet, wohingegen beim Blockparadigma (links) bei allen untersuchten zeitlichen Korrelationskoeffizienten Überschreitungen bis auf maximale Werte für den empirischen Fehler 1. Art von knapp zehn Prozent auftreten. Bei Anwendung des Blockpermutationsverfahrens kann das nominale Testniveau beim ereignisbezogenen Design für alle untersuchten Korrelationsstärken, beim Blockdesign bis einschließlich $\rho = 0,6$ approximativ eingehalten werden. Für $\rho = 0,8$ steigt der empirische Fehler 1. Art auf 6,4 Prozent an.

Die Ergebnisse der skaleninvarianten Version der Teststatistik gleichen denen der skalenabhängigen Form auch für den modifizierten PC-Test sehr stark und sind daher ebenso in Anhang C aufgeführt.

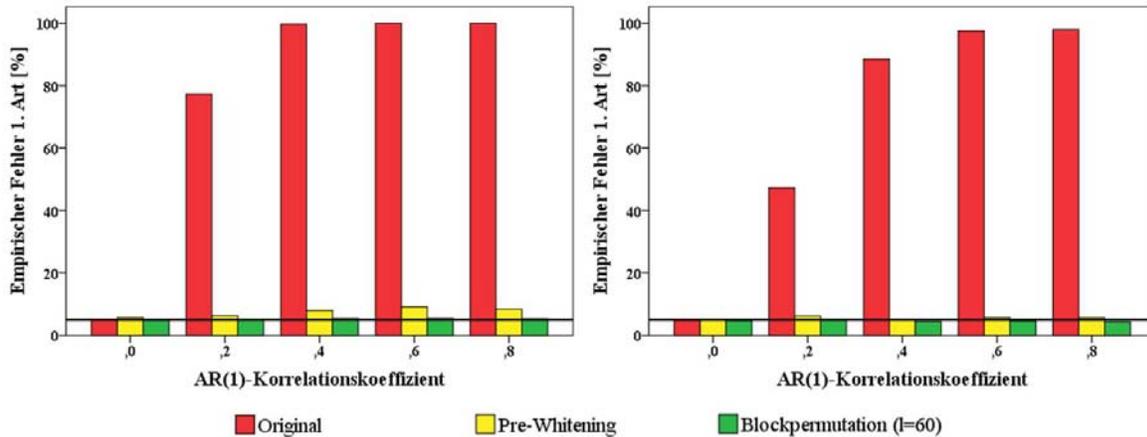


Abbildung 6.6.: Empirischer Fehler 1. Art des Tests auf Basis des Ähnlichkeitsmaßes in Form des Korrelationskoeffizienten nach Pearson bei simulierten fMRT-Daten für verschiedene Adjustierungsverfahren; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

Wird als Teststatistik ein Ähnlichkeitsmaß – hier der **Pearson'sche Korrelationskoeffizient** – verwendet, dessen Verteilung unter der Nullhypothese durch eine Permutation erzeugt wird, kann sowohl das Pre-Whitening als auch das Blockpermutationsverfahren zur Anpassung an die zeitliche Korrelation genutzt werden. Der entsprechende empirische Fehler 1. Art ist in Abbildung 6.6 dargestellt. Für den originalen Test ergibt sich für beide Designs bei hohen zeitlichen Korrelationen ein Rate für den empirischen Fehler 1. Art von 100 Prozent. Während durch die blockweise Permutation mit Random Shift sowohl beim Blockparadigma (links) als auch beim ereignisbezogenen Design (rechts) das nominale Testniveau für alle untersuchten Korrelationsstärken approximativ eingehalten werden kann, gilt dies auch bei dieser Teststatistik unter Verwendung des Pre-Whitenings nur für das ereignisbezogene Design. Für mittlere und hohe AR(1)-Korrelationskoeffizienten beträgt der empirische Fehler 1. Art hier noch 9,1 Prozent.

Der empirische Fehler 1. Art basierend auf einer Abstandsstatistik in Form des **quadratischen euklidischen Abstandes** ist in Abbildung 6.7 verbildlicht. Auch hier kann die

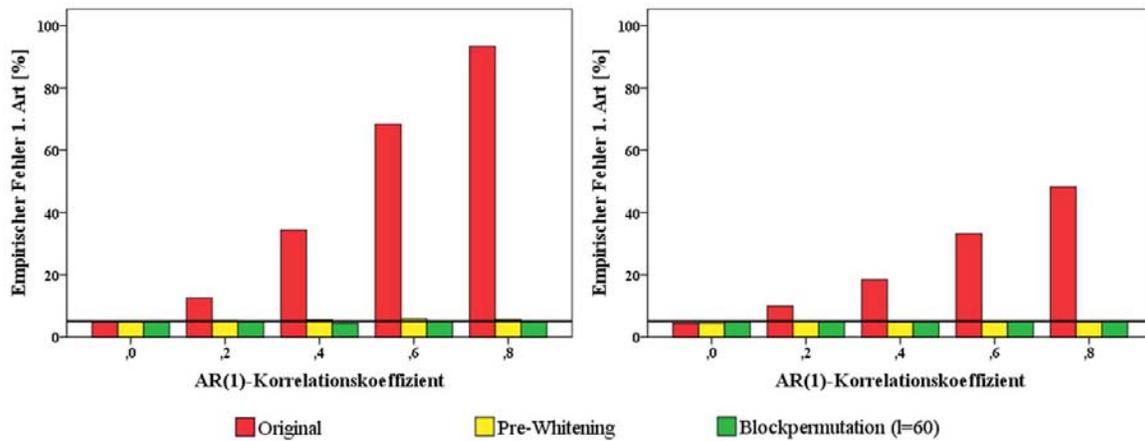


Abbildung 6.7.: Empirischer Fehler 1. Art des Tests auf Basis des Abstandsmaßes in Form des quadratischen euklidischen Abstandes bei simulierten fMRT-Daten für verschiedene Adjustierungsverfahren; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

Verletzung des nominalen Testniveaus durch die Autokorrelation der Stichprobenvektoren, die beim Blockdesign (links) bis auf etwa 93 Prozent und beim ereignisbezogenen Design (rechts) auf knapp 50 Prozent ansteigt, durch Pre-Whitening oder die blockweise Permutation mit Random Shift verringert werden. Im Gegensatz zum Ähnlichkeitsmaß wird das Testniveau hier für alle untersuchten Korrelationskoeffizienten approximativ eingehalten – sowohl durch die Blockpermutation als auch durch das parametrische Adjustierungsverfahren.

Zusammenfassung der bisherigen Ergebnisse:

Das nominale Testniveau wird für alle untersuchten multivariaten Teststatistiken mit wachsendem AR(1)-Korrelationskoeffizienten überschritten, wenn die Autokorrelation der Stichprobenvektoren unberücksichtigt bleibt. Für hohe Korrelationen (maximal $\rho = 0,8$ untersucht) können dabei mitunter sämtliche Nullhypothesen fälschlicherweise abgelehnt werden. Dies entspricht der Größenordnung des Fehlers 1. Art, wie sie auch für die verschiedenen Teststatistiken bei Durchführung der klassischen Permutation (vgl. Abschnitt 5.3) beobachtet werden kann.

Unter Verwendung der **Satterthwaite-Approximation** kann die Verletzung des nominalen Testniveaus für alle damit untersuchten Teststatistiken wie auch Korrelationsstärken behoben werden. Dabei sind nur geringe Unterschiede zwischen den beiden fMRT-Designs erkennbar. Allerdings wird dieses Adjustierungsverfahren nur auf die klassischen auf Hauptkomponenten basierenden stabilisierten multivariaten Teststatistiken angewendet. Zudem werden sämtliche Testverfahren ab mittleren AR(1)-Korrelationskoeffizienten in hohem Maße konservativ, so dass eine schlechte Güte für die Teststatistiken mit Satterthwaite-Approximation zu erwarten ist.

Das **Pre-Whitening** kann die Verletzung des nominalen Testniveaus bis zu mittleren

Korrelationsstärken in allen simulierten Fällen beheben. Für höhere Korrelationen ist dies beim ereignisbezogenen Design ebenso auf Basis aller untersuchten Teststatistiken erreichbar. Lediglich beim fMRT-Blockparadigma und hohen AR(1)-Korrelationen erzielt die Mehrzahl der Teststatistiken unter Pre-Whitening-Verwendung mit Werten für den empirischen Fehler 1. Art von bis zu zehn Prozent noch eine Überschreitung des nominalen Testniveaus. Dies gilt in ähnlichem Maß für die univariate t -Teststatistik.

Wie bereits in Abschnitt 5.3 dargestellt, kann die Verletzung des nominalen Testniveaus in Abhängigkeit von der Autokorrelationsstärke der Stichprobenvektoren bei der nicht-parametrischen Adjustierung der Testverfahren durch eine geeignete Auswahl der Blocklänge approximativ behoben werden. Allein bei den q -dimensionalen PC-Teststatistiken für ein fMRT-Blockdesign reicht die hier gewählte Blocklänge von 60 aufeinanderfolgenden Stichprobenvektoren nicht aus, wenn die simulierte Korrelation Werte von $\rho = 0,6$ oder höher annimmt. Sofern die zu untersuchende Messreihe länger ist, sollten daher entsprechend größere Blocklängen genutzt werden. Für alle anderen untersuchten Szenarien kann das nominale Testniveau anhand der **Blockpermutation mit Random Shift** bei einer Blocklänge von 60 Stichprobenvektoren approximativ eingehalten werden.

Die hier noch auftretenden Überschreitungen des nominalen Testniveaus für hohe zeitliche Korrelationen können durch **Kopplung** der Adjustierungsverfahren behoben werden. Bereits in Naujoks (2008) wird für einfache Zweistichprobendesigns gezeigt, dass durch Durchführung der blockweisen Permutation mit Random Shift nach Anwendung des Pre-Whitenings eine stärkere Verminderung der Verletzung des Fehlers 1. Art erzielt wird als durch die alleinige Anwendung eines der Verfahren. Dies soll jedoch im Rahmen dieser Arbeit nicht detailliert verfolgt werden und wird daher im Weiteren nur beispielgebend für das Blockdesign bei Verwendung der q -dimensionalen skalenabhängigen Principal Component Teststatistik verdeutlicht.

Design	Testverfahren	ρ				
		0,0	0,2	0,4	0,6	0,8
Block	Pre-Whitening+Blockpermutation	0,047	0,052	0,058	0,044	0,054
Event	Pre-Whitening+Blockpermutation	0,045	0,050	0,045	0,047	0,048

Tabelle 6.2.: Empirischer Fehler 1. Art des q -dimensionalen skalenabhängigen Principal Component Tests für simulierte fMRT-Daten bei Durchführung von Pre-Whitening und anschließender Blockpermutation mit Random Shift ($n = 250$, $p = 1024$, $l = 60$, 299 Permutationen, 2500 Wiederholungen)

Tabelle 6.2 zeigt den Fehler 1. Art der q -dimensionalen skalenabhängigen PC-Teststatistik bei Verwendung des Blockpermutationsverfahrens mit Random Shift (Blocklänge 60) bei zuvor durchgeführtem Pre-Whitening. Im Gegensatz zur Durchführung der einzelnen Adjustierungsverfahren (vgl. Abbildung 6.4) kann das nominale Testniveau durch Kopplung von Pre-Whitening und dem Blockpermutationsverfahren für beide fMRT-Paradigmen sowie alle untersuchten zeitlichen Korrelationsstärken approximativ

eingehalten werden.

Die Situationen mit sehr hohen zeitlichen Korrelationen sind somit auch anhand der multivariaten Teststatistiken bei gekoppelten Adjustierungsmethoden analysierbar.

Während sich die Satterthwaite-Approximation also in den Analysen zum Fehler 1. Art als konservativ herausstellt, erzielen das Pre-Whitening wie auch die Blockpermutation mit Random Shift größtenteils ein approximatives Einhalten des nominalen Testniveaus. Ausnahmen davon betreffen hohe zeitliche Korrelationen, die in realen fMRT-Daten jedoch kaum erreicht werden, für die aber durch Kopplung der beiden Verfahren ebenfalls ein approximatives Einhalten des Testniveaus erzielt wird. Generell sind die multivariaten Verfahren demnach unter Anwendung der parametrischen oder des nichtparametrischen Adjustierungsverfahrens für autokorrelierte Stichprobenvektoren auf fMRT-Daten anwendbar. Welches Verfahren dabei die höchste Güte und einen Vorteil gegenüber der univariaten Analyse aufweist, wird in Abschnitt 6.3 untersucht.

6.2.2. Ergebnisse der simulierten Kontrollscenarien

Vor Beurteilung der Güte der einzelnen Test- und Adjustierungsverfahren für simulierte fMRT-Daten mit der angenommenen Block-Compound-Symmetry-Struktur für räumliche Korrelationen im Sinne funktioneller Gebiete wird nachfolgend zunächst auf die in Abschnitt 6.1 erwähnten Kontrollscenarien eingegangen.

Simulation einer zufälligen räumlichen Abhängigkeitsstruktur

Alle multivariaten Teststatistiken und Anpassungsvarianten an die zeitliche Korrelationsstruktur werden im Folgenden am Beispiel von $\rho = 0,4$ einander gegenübergestellt, wobei sich die räumliche Korrelation der 1024 Voxel in jedem Simulationsschritt zufällig ergibt.

Abbildung 6.8 enthält die Werte des empirischen Fehlers 1. Art für die untersuchten multivariaten Teststatistiken und zeitlichen Adjustierungsmethoden für das Blockdesign (links) und das ereignisbezogene Design (rechts). Zusätzlich ist das Ergebnis der t -Tests unter Pre-Whitening aufgeführt, für die sich ein über alle Variablen gemittelter empirischer Fehler 1. Art von 5,92 Prozent beim fMRT-Blockdesign und 5,48 Prozent beim ereignisbezogenen Design ergibt. Für die unkorrigierten multivariaten Teststatistiken wird auch in diesem Simulationsszenario das nominale Testniveau bei allen Tests deutlich überschritten. Vielfach liegen die Werte der unkorrigierten Verfahren über den entsprechenden Ergebnissen bei einer räumlichen fMRT-Abhängigkeitsstruktur. Die Dimension der PC-Scores liegt hierbei im Mittel bei $\bar{q} = 63$ (60 bis 65) beim Blockdesign und bei $\bar{q} = 62$ (59 bis 65) beim ereignisbezogenen Design. Die Schätzung des zeitlichen Korrelationskoeffizienten fällt dabei aber mit $\hat{\rho} = 0,3891$ ($\text{RMSE}_{\hat{\rho}} = 0,0112$) beim Blockdesign und $\hat{\rho} = 0,3892$ ($\text{RMSE}_{\hat{\rho}} = 0,0111$) beim ereignisbezogenen Design ähnlich gut aus wie bei den Simulationen mit räumlicher fMRT-Struktur.

Auch bei diesem Kontrollscenario bewirken sämtliche Adjustierungsverfahren für zeitliche Korrelation bei allen Teststatistiken eine starke Verringerung der Verletzung des

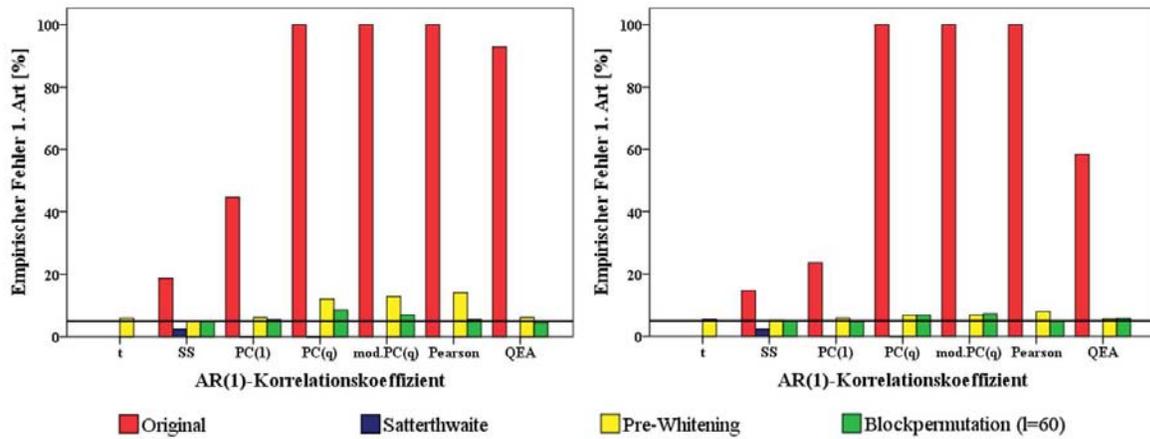


Abbildung 6.8.: Empirischer Fehler 1. Art der verschiedenen Adjustierungsverfahren und Tests bei simulierten fMRT-Daten mit zufälliger räumlicher Korrelation; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, $\rho = 0,4$, 299 Permutationen bzw. Rotationen, 2500 Wiederholungen)

nominalen Testniveaus. Während die Satterthwaite-Approximation auch hier sehr konservativ wird, kann das nominale Niveau jedoch vor allem beim Blockdesign unter Pre-Whitening nicht bei allen Testverfahren eingehalten werden. Der maximale empirische Fehler 1. Art liegt hierbei bei 14 Prozent. Beim Blockpermutationsverfahren wird das approximative Einhalten des Testniveaus lediglich beim klassischen wie auch modifizierten q -dimensionalen PC-Test verfehlt. Das zeigt aber, dass die adaptierten multivariaten Verfahren generell auch für diese stark überzeichnete räumliche Korrelationsstruktur anwendbar sind und die angenommene Struktur für funktionelle Gebiete somit keine übermäßige Begünstigung der Verfahren bewirkt.

Simulation einer verkürzten Zeitreihe

Die folgenden Ergebnisse zum empirischen Fehler 1. Art resultieren aus der Betrachtung der nach fMRT-Design simulierten Daten, jedoch bei einer stark verkürzten Messreihe von nur 50 Stichprobenvektoren.

Abbildung 6.9 zeigt den empirischen Fehler 1. Art bei verkürzter Messreihe für die einzelnen Teststatistiken und Adjustierungsmethoden bei einer zeitlichen Korrelationsstärke von $\rho = 0,4$ für das Blockdesign (links) sowie für das ereignisbezogene Design (rechts). Vor allem beim Blockdesign wird sichtbar, dass die Verletzung des nominalen Testniveaus hier nicht behoben werden kann. Selbst für den durch Pre-Whitening adjustierten t -Test ergibt sich in diesem Szenario ein mittlerer empirischer Fehler 1. Art von 13,6 Prozent.

Bei nur 50 aufeinanderfolgenden Messungen ist die Schätzung der zeitlichen Korrelation relativ schlecht: $\hat{\rho} = 0,3328$ ($\text{RMSE}_{\hat{\rho}} = 0,0697$) beim Blockdesign und $\hat{\rho} = 0,3433$ ($\text{RMSE}_{\hat{\rho}} = 0,0596$) beim ereignisbezogenen Design. Das erklärt bereits die anhaltende Überschreitungen des Testniveaus bei Verwendung der parametrischen Adjustierungs-

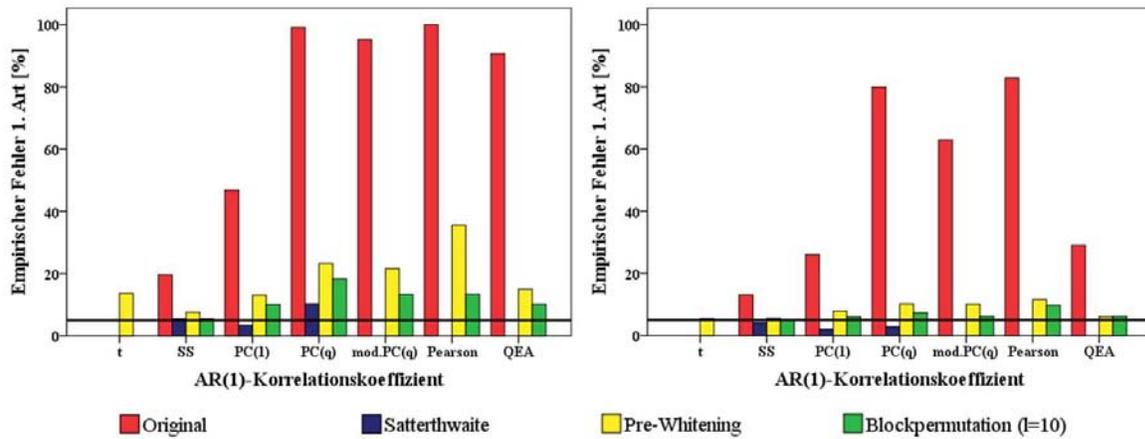


Abbildung 6.9.: Empirischer Fehler 1. Art der verschiedenen Adjustierungsverfahren und Tests bei simulierten fMRT-Daten einer verkürzten Zeitreihe; links Blockdesign, rechts ereignisbezogenes Design ($n = 50$, $p = 1024$, $\rho = 0,4$, 299 Permutationen bzw. Rotationen, 2500 Wiederholungen)

methoden. Auch für die Blockpermutation reicht diese sehr kurze Zeitreihe nicht aus, um approximativ austauschbare Stichprobenblöcke zu erzielen (vgl. Abschnitt 5.3). Bei der hier untersuchten Korrelation von $\rho = 0,4$ müssten dazu Blocklängen von etwa 40 aufeinanderfolgenden Stichprobenvektoren verwendet werden, was aber trotz der zufälligen Verschiebung bei nur $n = 50$ nicht möglich ist. Bei den dargestellten Ergebnissen erfolgt die blockweise Permutation bei einer Blocklänge von $l = 10$.

In etwas geringerem Maße gelten diese Restriktionen der Verfahren auch beim ereignisbezogenen Design. Hier kann das nominale Testniveau ebenso für den Großteil der Teststatistiken und Adjustierungsverfahren nicht approximativ eingehalten werden. Lediglich die Satterthwaite-Approximation bewirkt ein konservatives Testverhalten.

Diese extrem geringe Anzahl an Messungen wird jedoch in üblichen fMRT-Experimenten weit überschritten und sollte hier lediglich die Grenzen der Adjustierungsverfahren aufzeigen. Auf Güteuntersuchungen wird daher für dieses Szenario verzichtet.

Im Weiteren folgt nun die Betrachtung der einzelnen Teststatistiken und zeitlichen Adjustierungsmethoden unter der Alternativhypothese.

6.3. Gütesimulationen

6.3.1. Güteresultate der simulierten fMRT-Daten

Im Fokus der folgenden Güteuntersuchungen steht sowohl für das Blockdesign als auch für das ereignisbezogene Design die Simulation der Daten unter der Alternativhypothese, wobei jeweils die Differenz zweier Stimuli getestet wird (vgl. Abschnitt 6.1).

Es werden $p_1 = 128$ der 1024 Variablen unter der Alternativhypothese simuliert. Dabei handelt es sich um die ersten zwei der insgesamt 16 Variablenblöcke, die unabhängig von-

einander sind und innerhalb eines Blocks anhand einer Compound-Symmetry-Struktur miteinander korrelieren. Somit weisen die Variablen unter der Alternative keine einheitliche Abhängigkeitsstruktur auf. Das simulierte fMRT-Signal entspräche dabei beispielsweise einer jeweils gleichmäßigen Aktivierung in zwei verschiedenen funktionellen Gehirngebieten.

Die Ergebnisse der Güteuntersuchungen der verschiedenen multivariaten Teststatistiken und Adjustierungsverfahren für zeitliche Abhängigkeit werden in den folgenden Abschnitten miteinander verglichen. Um gleichzeitig eine Gegenüberstellung mit einer univariaten Testmethode zu realisieren, erfolgt die Simulation und Analyse unter der Alternative auch parallel je Variable.

Design	ρ				
	0,0	0,2	0,4	0,6	0,8
Blockdesign	0,984	0,908	0,709	0,476	0,340
ereignisbezogenes Design	0,931	0,818	0,716	0,734	0,948

Tabelle 6.3.: Empirische Güte der t -Teststatistiken für simulierte fMRT-Daten bei Durchführung von Pre-Whitening zum Nachweis mindestens einer signifikanten Aktivierung ($n = 250$, $p = 1024$, Bonferroni-korrigiert, 2500 Wiederholungen)

Die Güte der **univariaten t -Tests** bei Durchführung des Pre-Whitenings zur Adjustierung der zeitlichen Korrelation ist in Tabelle 6.3 dokumentiert. Sie spiegelt wider, wie häufig innerhalb der 1024 simulierten Voxel univariat mindestens eine signifikante Aktivierung unter Bonferroni-Korrektur nachgewiesen werden kann. Dieser Wert wird im Weiteren als Güte der “univariaten” Analyse bezeichnet. Dabei ist zu beachten, dass hier die Resultate aller Variablen berücksichtigt werden und dieser Test somit in gewisser Weise als multivariater Test verstanden werden kann.

Beim fMRT-Blockdesign ist gut erkennbar, dass die Güte der “univariaten” Tests mit steigender zeitlicher Korrelation abnimmt. Bei unkorrelierten Stichprobenvektoren wird hier in 98,4 Prozent der Simulationen mindestens eine Aktivierung nachgewiesen. Dieser Anteil sinkt für $\rho = 0,8$ auf 34 Prozent.

Beim ereignisbezogenen Design ist dieser Güteverlust bis zu mittleren zeitlichen Korrelationen ebenso erkennbar. Die “univariate” Güte sinkt hier von 93,1 auf 71,6 Prozent. Für noch höhere AR(1)-Korrelationskoeffizienten steigt der Anteil mindestens einer nachweisbaren Aktivierungen für dieses fMRT-Paradigma jedoch wieder auf knapp 95 Prozent an.

Zusätzlich zu diesem mit der Güte der multivariaten Tests gut vergleichbaren Wert wird für die univariaten Tests gezählt, wie oft für die einzelnen unter der Alternative simulierten Variablen im Mittel eine Signifikanz erreicht wird. Dies erfolgt anhand der unkorrigierten Irrtumswahrscheinlichkeit $\alpha = 0,05$, an der in der funktionellen Bildgebung häufig benutzten unkorrigierten Schranke von 0,1 Prozent sowie anhand der Bonferroni-Schranke von $\frac{0,05}{1024} = 4,883 \cdot 10^{-5}$.

Die Spannweite dieser variablenbezogenen Gütemaße ist für die $p_1 = 128$ Variablen unter

		ρ				
Design	Testniveau	0,0	0,2	0,4	0,6	0,8
Block	$\alpha = 0,05$	0,91/0,94	0,78/0,82	0,61/0,65	0,44/0,48	0,34/0,39
	$\alpha = 0,001$	0,51/0,55	0,29/0,33	0,14/0,17	0,07/0,09	0,04/0,07
	Bonferroni	0,22/0,26	0,09/0,12	0,03/0,05	0,01/0,02	0,00/0,01
Event	$\alpha = 0,05$	0,80/0,85	0,70/0,73	0,62/0,66	0,63/0,68	0,83/0,86
	$\alpha = 0,001$	0,32/0,36	0,20/0,24	0,14/0,18	0,15/0,19	0,34/0,39
	Bonferroni	0,10/0,14	0,05/0,07	0,03/0,05	0,03/0,05	0,11/0,15

Tabelle 6.4.: Empirische Güte der einzelnen t -Teststatistiken für simulierte fMRT-Daten bei Durchführung von Pre-Whitening für verschiedene Testniveaus der Einzeltests; Angabe in minimal und maximal erreichte Güte der 128 Variablen unter der Alternative (Min/Max, $n = 250$, $p = 1024$, 2500 Wiederholungen)

der Alternative für beide fMRT-Designs in Tabelle 6.4 in Abhängigkeit vom zeitlichen Korrelationskoeffizienten dargestellt. Für beide fMRT-Paradigmen ist deutlich erkennbar, dass die minimal und maximal erreichbare Güte in den einzelnen Variablen sich in Abhängigkeit vom Niveau für die einzelnen Tests stark unterscheidet.

Während beim Blockdesign für unkorrelierte Daten bei unadjustiertem Testniveau jede Variable unter der Alternative im Mittel zu über 90 Prozent als aktiviert erkannt wird, sinkt dieser Anteil bei der in fMRT-Analysen häufig verwendeten Signifikanzschranke von 0,1 Prozent bereits auf 51 bis 55 Prozent und unter Bonferroni-Korrektur schließlich auf etwa ein Viertel. Die Gütergebnisse verringern sich auch für die einzelnen Tests mit steigendem zeitlichen Korrelationskoeffizienten, so dass unter FWER-Einhaltung durch Bonferroni die Güte der einzelnen Tests für $\rho = 0,8$ im Mittel unter einem Prozent liegt. Für das ereignisbezogene Design sinkt der Anteil signifikant erkannter Variablen ebenso mit kleiner werdender Signifikanzschranke. Doch entsprechend der Güte der “univariaten” Tests in Tabelle 6.3 steigt die Güte der Einzeltests für dieses fMRT-Paradigma bei hohen zeitlichen Korrelationen wieder an. Selbst unter Bonferroni-Korrektur wird die Aktivierung in den Variablen unter der Alternative bei $\rho = 0,8$ zu elf bis 15 Prozent gefunden.

Bei hohen zeitlichen Korrelationen und zugrunde liegendem ereignisbezogenem Design ist das Verhalten der Tests, unter Verwendung des Pre-Whitenings wieder ansteigende Gütergebnisse zu erzielen, im Weiteren auch bei den multivariaten Verfahren zu beobachten. Dies ist speziell in diesem Simulationsszenario damit zu begründen, dass das Eliminieren der hohen zeitlichen Korrelation aus den Daten durch Pre-Whitening parallel auch für das Design erfolgt (vgl. Abschnitt 4.1.2). Beim ereignisbezogenen Design sind die Stimuli, gerade wenn sie in kurzen Intervallen erfolgen, durch die Faltung mit der hämodynamischen Responsefunktion teilweise sehr verwischt. Das bedeutet, das Baseline-niveau zwischen aufeinanderfolgenden Stimuli wird unter Umständen nicht erreicht, da sich der nächste Stimulus direkt anschließt. Je höher nun die Glättung ist, die durch

das Pre-Whitening aus Daten und Design entfernt wird, desto besser sind die einzelnen Stimuli auch im Design wieder voneinander distanziert erkennbar, womit folglich auch die Güte der Testverfahren ansteigt.

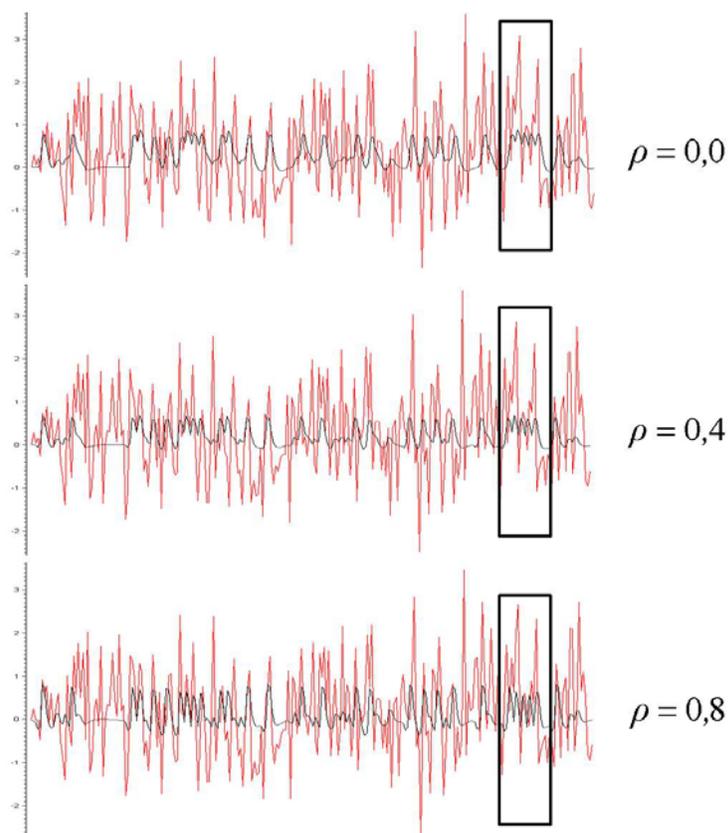


Abbildung 6.10.: Darstellung eines simulierter fMRT-Datenvektors (rot) und des interessierenden Kontrasts (schwarz) beim ereignisbezogenen Design nach Pre-Whitening für verschiedene zeitliche Korrelationskoeffizienten ($n = 250$)

Dieses Verhalten ist anhand eines Beispielvektors in Abbildung 6.10 dargestellt. Die rote Linie stellt das simulierte Signal für ein Voxel nach Durchführung des Pre-Whitenings dar. Die schwarze Linie entspricht dem Design, genauer dem aus dem Paradigma durch Faltung und Kontrastbildung gebildeten Vektor, der auf das simulierte Signal addiert worden ist, um die entsprechende Variable unter der Alternative zu simulieren – und damit dem zu untersuchenden Kontrast (vgl. Abschnitt 6.1). Auch dieser Kontrast ist anhand der aus den Daten geschätzten zeitlichen Korrelation “pre-whitened”, um approximativ ein Modell mit unabhängigen Stichprobenvektoren zu erreichen.

Der eingerahmte Teil der Zeitreihe in Abbildung 6.10 zeigt einen Bereich schnell aufeinanderfolgender Stimuli, an dem gut erkennbar ist, wie sich das Pre-Whitening auf die Differenzierung der Stimuli voneinander auswirkt. Bei $\rho = 0,0$ wird formal zwar auch ein Pre-Whitening mit geschätztem Korrelationskoeffizienten vorgenommen, dabei bleibt der durch Faltung mit der hämodynamischen Responsefunktion erhaltene Stimulusverlauf für den interessierenden Kontrast jedoch nahezu unverändert. Bei einer

mittleren zeitlichen Korrelationsstärke zeichnet sich bereits eine leichte Veränderung des Kontrasts ab – er ist weniger geglättet. Diese “Ent-Glättung” wird aber erst bei hohen zeitlichen Korrelationen richtig deutlich. Am Beispiel des hervorgehobenen Bereichs für $\rho = 0,8$ sieht man, dass der Kontrast nach Durchführung des Pre-Whitenings zwischen den einzelnen Stimuli stärker absinkt und somit besser zum entkorrelierten Signal passt. Deshalb steigt die Testgüte in diesem Design bei Durchführung des Pre-Whitenings für hohe zeitliche Korrelationen an.

Die folgenden Bilder zeigen für je eine multivariate Teststatistik für die verschiedenen Adjustierungsverfahren die empirische Güte, also den prozentualen Anteil richtig abgelehnter Nullhypothesen für das fMRT-Blockdesign auf der linken Seite und rechts für das ereignisbezogene Paradigma. Da der empirische Fehler 1. Art für die Mehrheit der Adjustierungsverfahren in allen Parametereinstellungen approximativ dem nominalen Testniveau entspricht, werden sämtliche Testergebnisse für alle untersuchten Korrelationsstärken dargestellt. Jene Stellen, an denen der Fehler 1. Art vom nominalen Testniveau abweicht, werden im Text entsprechend beachtet.

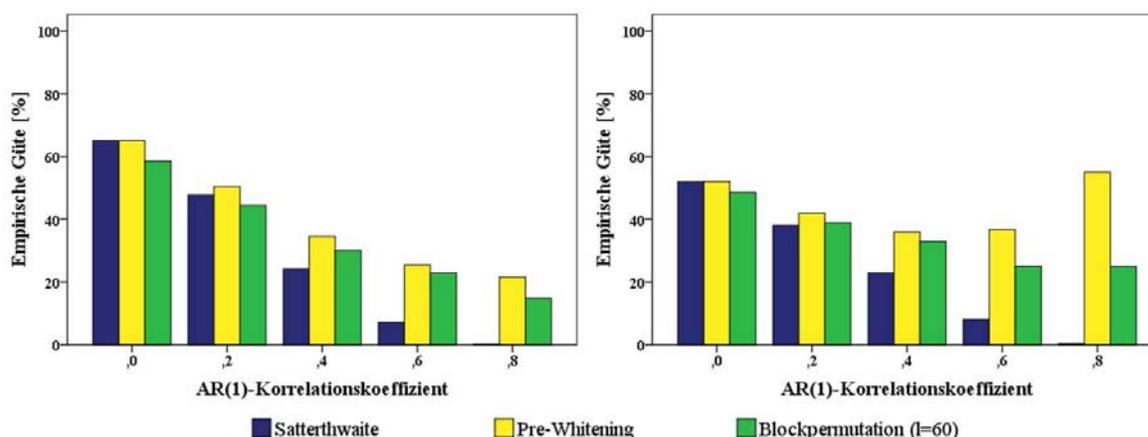


Abbildung 6.11.: Empirische Güte der verschiedenen Adjustierungsverfahren des Standardisierten Summentests bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 128 Variablen unter der Alternative, 299 Permutationen, 2500 Wiederholungen)

Abbildung 6.11 zeigt die Güte der verschiedenen Adjustierungsverfahren für den **Standardisierten Summentest** in Abhängigkeit vom AR(1)-Korrelationskoeffizienten. Für das fMRT-Blockdesign (links) erreicht die Standardisierte Summenteststatistik bei unkorrelierten Stichprobenvektoren ($\rho = 0,0$) sowohl unter Satterthwaite-Approximation, als auch beim Pre-Whitening eine Güte von etwa 65 Prozent. Bei Verwendung der Blockpermutation mit Random Shift werden die richtig-positiven Variablen mit 58,6 Prozent zu einem etwas geringeren Anteil erkannt. Es ist erkennbar, dass die Güte bei allen Adjustierungsmethoden mit steigender zeitlicher Korrelation unter sonst gleichbleibenden Parametereinstellungen erwartungsgemäß sinkt.

Zudem verliert die Satterthwaite-Approximation mit steigender zeitlicher Abhängigkeit sehr stark an Güte, so dass schließlich für den höchsten untersuchten AR(1)-Korrelationskoeffizienten ($\rho = 0,8$) keinerlei Aktivierung in den Variablen gefunden werden kann. Im Vergleich dazu ist der Güteverlust durch steigende Abhängigkeit der Stichprobenvektoren beim Pre-Whitening wie auch bei der Blockpermutation mit Random Shift vergleichsweise gering. Der Anteil richtig-positiver Ergebnisse liegt dabei beim nichtparametrischen Adjustierungsverfahren stets etwa fünf Prozent unter denen bei Pre-Whitening. Für die Satterthwaite-Approximation der Standardisierten Summenteststatistik sehen die Simulationsergebnisse bei zugrunde liegendem ereignisbezogenem fMRT-Paradigma (rechts) größtenteils ähnlich aus. Die Güte des Verfahrens sinkt mit steigender zeitlicher Korrelation monoton von 52 ($\rho = 0,0$) auf 0,7 Prozent ($\rho = 0,8$) ab. Bei nichtmetrischer Adaption der Standardisierten Summenteststatistik für zeitliche Korrelation werden auch für das ereignisbezogene Design die jeweils zweitbesten Gütergebnisse der Teststatistik erzielt. Der Wert sinkt hier von 48,5 Prozent erkannter richtig-positiver Ergebnisse bei $\rho = 0,0$ auf 25 Prozent bei $\rho = 0,8$. Beim Pre-Whitening ist hingegen deutlich sichtbar, dass auch die Güte der Standardisierten Summenteststatistik bei dieser Adjustierungsmethode mit wachsender Korrelation nicht monoton fällt. Der Wert sinkt zwar zunächst von 52 Prozent bei $\rho = 0,0$ auf 36 Prozent bei $\rho = 0,4$, steigt dann aber schließlich für die größte untersuchte Korrelationsstärke wieder auf 55 Prozent an. Dies entspricht dem Verhalten der Güte der “univariaten” Tests beim ereignisbezogenen Design wie auch den folgenden multivariaten Teststatistiken bei Verwendung des Pre-Whitenings. Mit diesen Ergebnissen liegt die Güte des Standardisierten Summentests jedoch für alle Adjustierungsverfahren wie auch Parametereinstellungen unter der des “univariaten” Verfahrens (vgl. Tabelle 6.3).

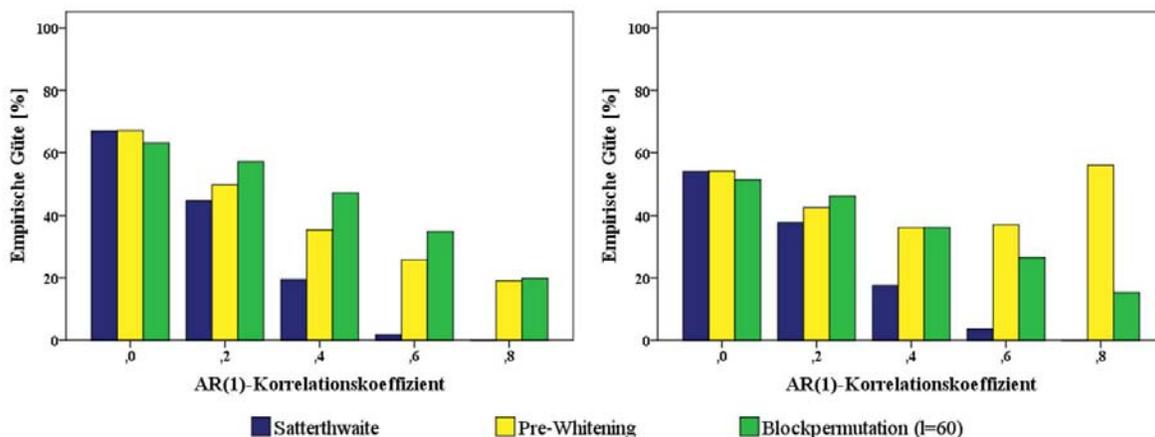


Abbildung 6.12.: Empirische Güte der verschiedenen Adjustierungsverfahren des eindimensionalen skalenabhängigen Principal Component Tests bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 128 Variablen unter der Alternative, 299 Permutationen, 2500 Wiederholungen)

Abbildung 6.12 veranschaulicht die Güte der **eindimensionalen skalenabhängigen Principal Component Teststatistik** für die verschiedenen Adaptionsverfahren für zeitliche Korrelation. Für beide fMRT-Designs liegen die Ergebnisse etwa in derselben Größenordnung wie die für den Standardisierten Summentest. Auch hier verringert sich die Güte erwartungsgemäß mit steigender zeitlicher Korrelation. Eine Ausnahme bildet dabei wiederum das Pre-Whitening bei hohen Korrelationen für das ereignisbezogene Design (rechts).

Die Satterthwaite-Approximation verliert für beide fMRT-Designs am schnellsten an Güte und scheint deshalb für mittlere und hohe zeitliche Korrelationen zur Erkennung der Aktivierung nicht nutzbar zu sein.

Der Unterschied bei Verwendung dieser Statistik im Vergleich zur Standardisierten Summenteststatistik ist jedoch, dass bei geringen und mittleren zeitlichen Korrelationen beim ereignisbezogenen Design und generell beim Blockdesign das Blockpermutationsverfahren eine etwas höhere Güte aufweist als die Verwendung des Pre-Whitenings. Dennoch liegen auch hier sämtliche Ergebnisse unter denen der “univariaten” Tests.

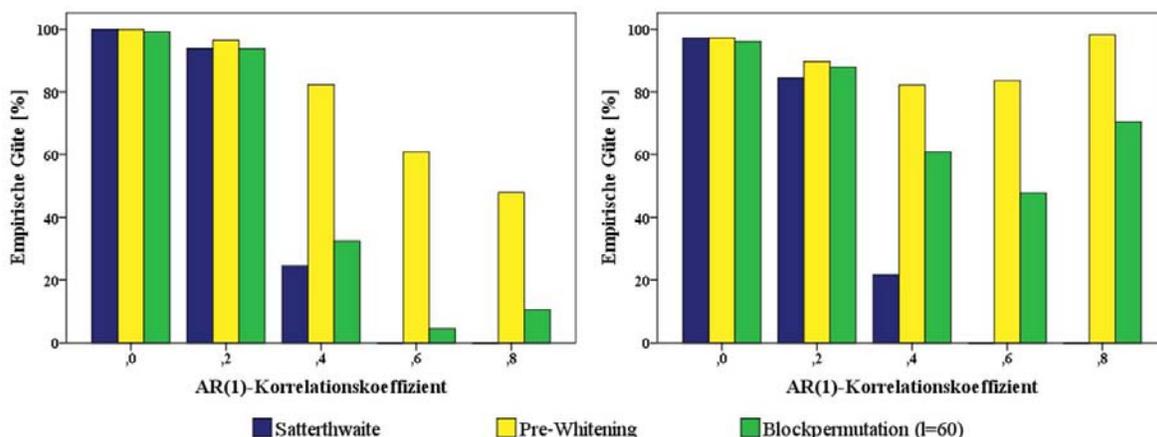


Abbildung 6.13.: Empirische Güte der verschiedenen Adjustierungsverfahren des q -dimensionalen skalenabhängigen Principal Component Tests bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 128 Variablen unter der Alternative, 299 Permutationen, 2500 Wiederholungen)

Die Güte der verschiedenen Adjustierungsverfahren bei Nutzung der **q -dimensionalen skalenabhängigen Principal Component Teststatistik** wird für beide fMRT-Paradigmen in Abbildung 6.13 veranschaulicht. Erwartungsgemäß sind die Ergebnisse besser als beim eben beschriebenen eindimensionalen Pendant, da die multivariaten Scores mehr Varianz der originalen Beobachtungen widerspiegeln als der erste Score allein und dieser Score schwer zu optimieren ist. Diese zusätzliche Information führt sowohl beim Blockdesign (links) als auch beim ereignisbezogenen Paradigma (rechts) zu einem größeren Anteil richtig-positiver Ergebnisse.

Beim Blockdesign wird die falsche Nullhypothese bei Unabhängigkeit der Stichproben-

vektoren von allen drei Adaptionstrategien zu über 99 Prozent abgelehnt und übersteigt damit die “univariaten” Güteresultate. Dieser Anteil sinkt wieder erwartungsgemäß mit steigender Korrelationsstärke ρ . Hierbei wird bei geringen Korrelationen ($\rho = 0,2$) noch bei allen drei Verfahren in mehr als 93 Prozent der Simulationen ein richtig-positives Ergebnis erzielt. Ab $\rho = 0,4$ nehmen die Resultate der Satterthwaite-Approximation wie auch die bei Blockpermutation mit Random Shift stark ab, während durch Pre-Whitening die simulierte Aktivierung bei $\rho = 0,8$ noch etwa zur Hälfte erkannt wird. Zumindest bei Verwendung des Pre-Whitenings liegen die Güteresultate für die q -dimensionale PC-Teststatistik damit in allen Fällen über den “univariaten” Ergebnissen. Dabei muss aber beachtet werden, dass der empirische Fehler 1. Art sowohl bei Verwendung des Pre-Whitenings wie auch für die nichtmetrische Adjustierung für diese Teststatistik bei höheren zeitlichen Korrelationen leicht oberhalb des nominalen Testniveaus liegt. Ein Teil des Anstiegs der Güte für diese Situation bei $\rho = 0,8$ erklärt sich somit durch das leicht liberale Verhalten der Tests.

Beim ereignisbezogenen Design liegt die Güte bei Pre-Whitening-Adjustierung für dieses Simulationsszenario stets oberhalb von 83 Prozent. Die Satterthwaite-Approximation funktioniert nur für schwache zeitliche Korrelationen gut. Bei $\rho = 0,4$ wird die Aktivierung noch in knapp 23 Prozent der Simulationen erkannt, für $\rho = 0,6$ und $\rho = 0,8$ kann kein richtig-positives Ergebnis erzielt werden. Bei der nichtparametrischen Adaptionmethode für zeitliche Autokorrelation sinkt die Güte zwar bis $\rho = 0,6$ auf knapp 50 Prozent ab, steigt dann jedoch analog zum Pre-Whitening für $\rho = 0,8$ wieder auf über 70 Prozent an. Im Gegensatz zur eindimensionalen PC-Teststatistik zeigen die höheren Scores jedoch keinen Vorteil des nichtmetrischen Verfahrens gegenüber dem Pre-Whitening. Die durch Pre-Whitening adjustierten Testergebnisse sind auch für das ereignisbezogene Design stets deutlich oberhalb der “univariaten” Resultate.

Bei Verwendung einer **skaleninvarianten Principal Component Teststatistik** werden bei allen drei Adaptionsverfahren für korrelierte Stichprobenvektoren Ergebnisse für die Güte erzielt, die nur geringfügig von denen mit skalenabhängiger Teststatistik abweichen. Dies gilt sowohl für den ein- und mehrdimensionalen Fall der zugrunde liegenden Scores wie auch für beide fMRT-Designs.

Der Grund dafür ist, dass die (simulierten) fMRT-Daten in ihrer generellen physikalischen Größenordnung für unterschiedliche Gehirnregionen wie auch Aktivierungszustände gleich sind. Die Verwendung der skalenunabhängigen Teststatistik ist hier somit nicht erforderlich und bietet keinen Vorteil gegenüber der skalenabhängigen Variante. Allerdings ist auch keine Überlegenheit der skalenabhängigen Verfahren erkennbar, deren Ergebnisse nur minimal über denen der skaleninvarianten Versionen liegen. Die Ergebnisse der entsprechenden Gütesimulationen sind daher hier nicht separat dargestellt, sondern in Anhang D tabellarisch aufgeführt.

Abbildung 6.14 stellt die Ergebnisse der Simulationen unter der Alternative für die **modifizierte PC-Teststatistik** dar. Auch hier unterscheiden sich die Ergebnisse der skalenabhängigen kaum von denen der skaleninvarianten Version, die daher ebenso in Anhang D tabellarisch aufgeführt sind.

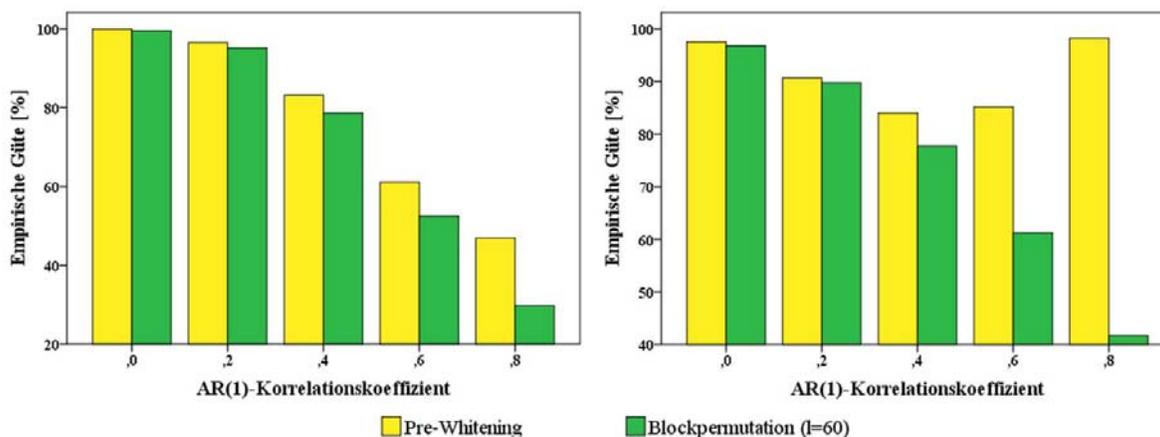


Abbildung 6.14.: Empirische Güte der verschiedenen Adjustierungsverfahren des modifizierten q -dimensionalen skalenabhängigen Principal Component Tests bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 128 Variablen unter der Alternative, 299 Rotationen bzw. Permutationen, 2500 Wiederholungen)

Die Güte der modifizierten PC-Teststatistik unter Pre-Whitening entspricht für beide fMRT-Designs in etwa der Größenordnung, die der klassische q -dimensionale PC-Test unter Pre-Whitening erreicht. Generell gibt es lediglich eine Verbesserung von etwa zwei Prozent. Damit weist der Test aber für alle untersuchten zeitlichen Korrelationsstärken eine höhere Güte auf als die “univariate” Analyse.

Bei zugrunde liegendem Blockdesign sinkt die Güte des Tests unter Pre-Whitening mit steigendem ρ von 99,9 auf 47 Prozent. Allerdings ist hierbei zu beachten, dass der empirische Fehler 1. Art dieses Tests das nominale Testniveau für mittlere und hohe Korrelationen um etwa drei Prozent überschreitet.

Beim ereignisbezogenen Design erzielt der modifizierte q -dimensionale PC-Test unter Pre-Whitening in allen untersuchten Fällen Werte über 84 Prozent, sinkt zunächst mit wachsendem ρ und steigt analog zu den anderen Verfahren bei Pre-Whitening-Adjustierung für höhere zeitliche Korrelationen leicht wieder an.

Die Anwendung der blockweisen Permutation mit zufälliger Verschiebung erzielt etwas geringere Ergebnisse als die Pre-Whitening-Adjustierung bei Verwendung eines Rotationsverfahrens. Die Resultate überschreiten die entsprechenden Ergebnisse bei “univariater” Analyse beim Blockdesign noch bis zur Korrelationsstärke $\rho = 0,6$, beim ereignisbezogenen Design bis zu mittleren Korrelationen von $\rho = 0,4$.

Abbildung 6.15 stellt die Ergebnisse der Güteuntersuchungen bei Verwendung einer Ähnlichkeitsstatistik, dem **Pearson’schen Korrelationskoeffizienten**, für die simulierten AR(1)-Korrelationsstärken dar.

Beim Blockdesign (links) erreicht die Güte bei den hier simulierten Parametereinstellungen für unkorrelierte Daten etwa 64 Prozent bei Verwendung des Pre-Whitenings, und knapp 30 Prozent beim Blockpermutationsverfahren. Der Anteil richtig-positiver

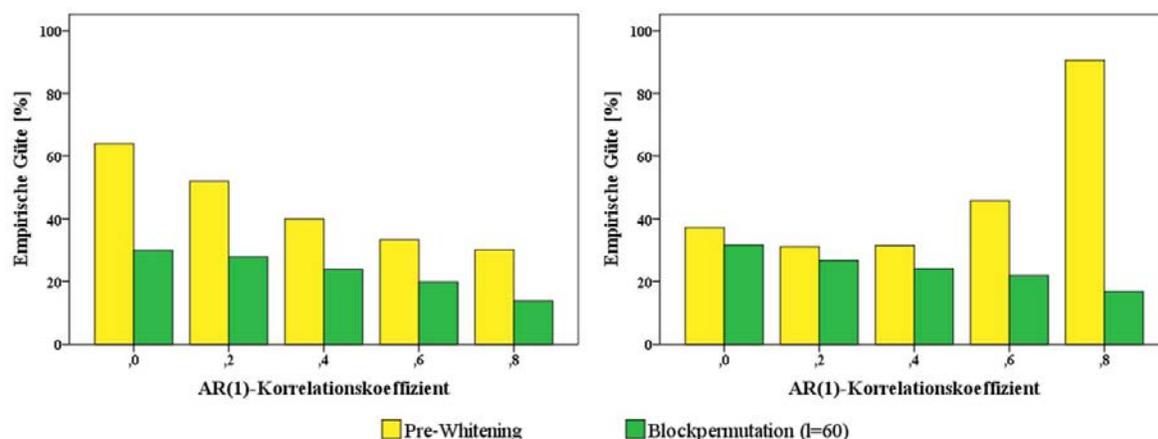


Abbildung 6.15.: Empirische Güte der verschiedenen Adjustierungsverfahren des Tests auf Basis des Pearson'schen Korrelationskoeffizienten bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 128 Variablen unter der Alternative, 299 Permutationen, 2500 Wiederholungen)

Ergebnisse verringert sich erwartungsgemäß mit wachsendem ρ , wobei die Ergebnisse bei blockweiser Permutation für $\rho = 0,8$ auf 13,8 Prozent, die des Pre-Whitenings auf 30 Prozent fallen. Die Ergebnisse bei Pre-Whitening sind jedoch für mittlere und hohe zeitliche Korrelationen etwas überschätzt, da das Verfahren bei Verwendung der Ähnlichkeitsstatistik mit bis zu 9,1 Prozent falsch-positiver Resultate leicht liberal ist.

Beim ereignisbezogenen Design (rechts) ist der Vorteil des Pre-Whitenings gegenüber dem nichtmetrischen Verfahren bis zu mittleren Korrelationsstärken wesentlich geringer als beim Blockdesign (wobei das nominale Testniveau hier bei beiden Verfahren approximativ eingehalten wird). Während die Ergebnisse der Gütesimulation für die blockweise Permutation mit Random Shift für $\rho = 0,6$ und $\rho = 0,8$ weiterhin abfallen, steigen die Werte beim Pre-Whitening wieder an und übersteigen schließlich den Anteil richtig-positiver Resultate bei Unabhängigkeit um das Doppelte. Das Pre-Whitening am Design wirkt sich bei Verwendung des Pearson'schen Korrelationskoeffizienten als Teststatistik damit am stärksten aus.

Dennoch reichen bei keinem der Adjustierungsverfahren die Testergebnisse basierend auf dem Test des Pearson'schen Korrelationskoeffizienten an die Gütergebnisse bei "univariater" Analyse heran.

Abbildung 6.16 veranschaulicht die Gütergebnisse des Tests auf Basis einer Abstandsstatistik bei Pre-Whitening und Blockpermutationsverfahren für beide fMRT-Designs. Bei Nutzung des **quadratischen euklidischen Abstandes** erzielen beide Adaptionsmethoden bei unabhängigen Stichprobenvektoren und zugrunde liegendem Blockdesign (links) nahezu gleiche Resultate von 52 Prozent. Mit wachsendem ρ verringert sich die Güte bei Pre-Whitening jedoch stärker als die des Blockpermutationsverfahrens.

Ähnlich verhält es sich beim ereignisbezogenen Paradigma (rechts). Der Vorteil der Blockpermutation gegenüber dem Pre-Whitening reicht hier jedoch nur bis zu mittleren

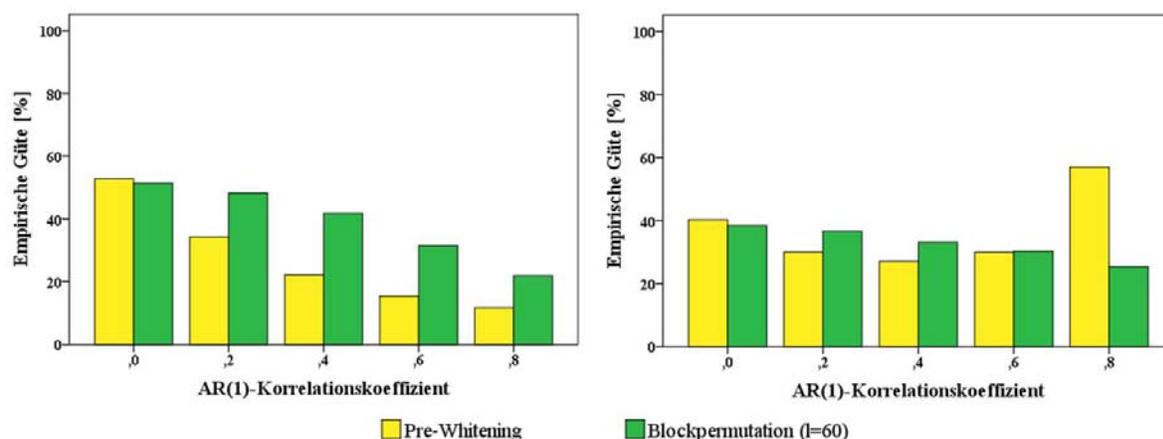


Abbildung 6.16.: Empirische Güte der verschiedenen Adjustierungsverfahren des Tests auf Basis des quadratischen euklidischen Abstandes bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 128 Variablen unter der Alternative, 299 Permutationen, 2500 Wiederholungen)

Korrelationskoeffizienten. Danach steigt auch bei dieser Teststatistik die Güte bei einer Pre-Whitening-Adjustierung wieder an, während die des Blockpermutationsverfahrens fällt.

Im Vergleich zur “univariaten” Analyse ist jedoch bei dieser Gütesimulation auch für die Abstandsstatistik kein Vorteil erkennbar.

Zusammenfassung der bisherigen Ergebnisse:

Für die hier simulierte Alternativhypothese liefern die multivariaten Teststatistiken in Abhängigkeit von der zeitlichen Korrelationsstärke stark unterschiedliche Güteresultate. Generell verringert sich die Güte erwartungsgemäß mit zunehmender Stärke der Autokorrelation der Stichprobenvektoren bei Verwendung der Satterthwaite-Approximation sowie der nichtmetrischen Blockpermutation mit zufälliger Verschiebung. Bei Adaption der Teststatistik an die zeitliche Korrelation durch das Pre-Whitening-Verfahren sinkt die Güte, ebenso beim Blockdesign mit wachsender Korrelation; beim ereignisbezogenen Design geschieht dies jedoch nur bis zu mittleren Korrelationsstärken. Für $\rho = 0,6$ und $\rho = 0,8$ steigt die Güte hingegen aufgrund des analog im Design durchgeführten Pre-Whitenings teilweise wieder stark an (vgl. Abbildung 6.10).

Von den einzelnen multivariaten Teststatistiken sind in diesem Simulationsszenario nur die mehrdimensionalen PC-Teststatistiken sowie deren modifizierte Form dem “univariaten” Verfahren überlegen.

Um die einzelnen Teststatistiken und ihre Adaptionrealisierungen miteinander zu vergleichen, zeigt Abbildung 6.17 die Ergebnisse der verschiedenen Verfahren für $\rho = 0,4$. Das Güteresultat der “univariaten” Analyse, das den Nachweis mindestens einer univariaten Aktivierung unter Bonferroni-Korrektur beschreibt, ist dabei ebenfalls abgebildet. Auch hier ist links das Blockparadigma und rechts das ereignisbezogene Design darge-

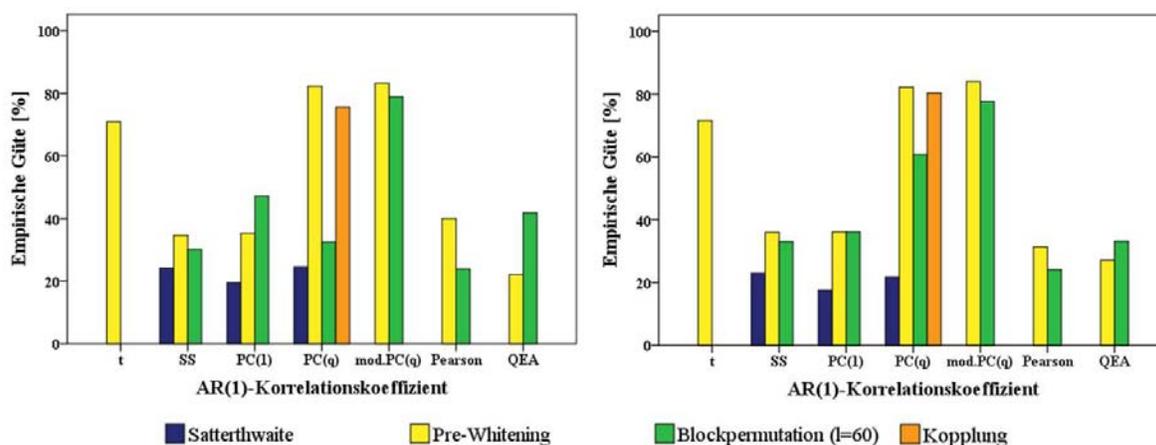


Abbildung 6.17.: Empirische Güte der verschiedenen Adjustierungsverfahren und Tests bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 128 Variablen unter der Alternative, $\rho = 0,4$, 299 Permutationen bzw. Rotationen, 2500 Wiederholungen)

stellt. Zusätzlich zu den untersuchten Adjustierungsverfahren für zeitliche Korrelation wird am Beispiel des q -dimensionalen PC-Tests auch die Kopplung von Pre-Whitening und dem Blockpermutationsverfahren betrachtet (vgl. Tabelle 6.2).

Bei zugrunde liegendem fMRT-Blockdesign (links) erreicht die Verwendung der modifizierten q -dimensionalen PC-Teststatistik unter Pre-Whitening-Adjustierung die höchste Güte von über 83 Prozent im Vergleich aller Verfahren. Der nächstbeste Wert von 82,3 Prozent wird durch die klassische q -dimensionale skalenabhängige PC-Teststatistik unter Pre-Whitening erzielt. Danach folgt das Ergebnis der modifizierten q -dimensionalen PC-Teststatistik bei blockweiser Permutation mit 78,9 Prozent Güte und das für Kopplung von Pre-Whitening und Blockpermutation bei Anwendung auf die klassische q -dimensionale PC-Teststatistik mit 75,6 Prozent Güte. Die Kopplung stellt somit keinen großen Güteverlust gegenüber dem allein durchgeführten Pre-Whitening dar, garantiert aber für alle untersuchten Korrelationsstärken die approximative Einhaltung des Testniveaus. Doch nur die bisher genannten Testmethoden sind der “univariaten” Testgüte von 70,9 Prozent überlegen. Im Weiteren folgen die eindimensionale PC-Teststatistik wie auch die Abstandsstatistik jeweils bei Durchführung des Blockpermutationsverfahrens mit 47 und 42 Prozent Güte. Die Satterthwaite-Approximation schneidet im Vergleich zu den anderen Methoden zur Adjustierung der zeitlichen Korrelation bei den jeweiligen Teststatistiken am schlechtesten ab.

Die Ergebnisse beim ereignisbezogenen Design sind denen des Blockdesigns sehr ähnlich. Auch hier weisen die q -dimensionalen Principal Component Teststatistiken bei Pre-Whitening und Blockpermutation wie auch Kopplung beider Adjustierungsverfahren mit Werten für die Güte von 77,7 bis 84 Prozent die besten Resultate auf und liegen über der entsprechenden Güte der “univariaten” Tests von 71,6 Prozent. Nach der Pre-Whitening-Version des q -dimensionalen PC-Tests (60 Prozent Güte) liegen Pre-Whitening und Blockpermutationsadjustierung für den eindimensionalen PC-Test und den Standardi-

sierten Summentest sowie die Pre-Whitening-adjustierte Ähnlichkeitsstatistik und der quadratische euklidische Abstand bei Verwendung der Blockpermutation mit Random Shift mit Werten zwischen 33 und 36 relativ nah beieinander.

Für die hier simulierten fMRT-Daten sind demnach vor allem die Teststatistiken vorteilhaft, die auf den q -dimensionalen Scores beruhen. Dabei erweist sich zumeist das Pre-Whitening zur Adaption an die zeitliche Korrelation der Daten als gütestärkstes Adjustierungsverfahren. Das ist nicht überraschend, da für die simulierten Daten die angenommene AR(1)-Korrelationsstruktur tatsächlich der zugrunde liegenden Struktur entspricht. Der hier vor allem bei der klassischen q -dimensionalen PC-Teststatistik auftretende Güteverlust bei Verwendung des Blockpermutationsverfahrens im Vergleich zum Pre-Whitening kann daher bei Anwendung auf reale Daten durchaus schwächer ausfallen. Lediglich die Satterthwaite-Approximation scheint den beiden anderen Verfahren gegenüber bei der hier simulierten Datenstruktur unterlegen zu sein.

6.3.2. Güteresultate der simulierten Kontrollszenarien

Verminderte Anzahl von Variablen unter der Alternative

Die Güteresultate der Simulationen, bei denen bei ansonsten gleichen Parametereinstellungen statt der 128 nur die Hälfte dieser Variablen unter der Alternativhypothese steht, werden in Abbildung 6.18 am Beispiel der AR(1)-Korrelationsstärke $\rho = 0,4$ für beide fMRT-Designs gegenübergestellt.

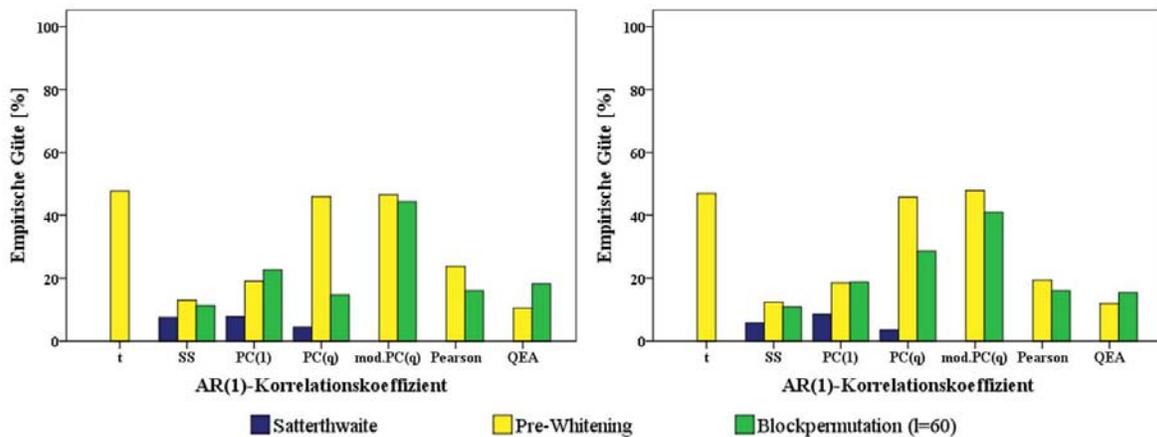


Abbildung 6.18.: Empirische Güte der verschiedenen Adjustierungsverfahren und Tests bei simulierten fMRT-Daten; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 64 Variablen unter der Alternative, $\rho = 0,4$, 299 Permutationen bzw. Rotationen, 2500 Wiederholungen)

Hierbei führt die geringere Anzahl der effektbehafteten Variablen bei allen Teststatistiken zur Verminderung der Güte, wobei sich das Verhalten der einzelnen Teststatistiken und verschiedenen Adjustierungsverfahren zueinander nicht grundsätzlich ändert. Auch

die Bestimmung der Scoredimension ist hier vor allem durch die räumliche Korrelationsstruktur in Form der 16 simulierten Gebiete bedingt und wird nicht maßgeblich durch die Anzahl der Blöcke unter der Alternative beeinflusst.

Analog zur vorhergehenden Güteuntersuchung weisen auch bei der Simulation nur eines Variablenblocks unter der Alternativhypothese der modifizierte wie auch der klassische mehrdimensionale Principal Component Test bei Pre-Whitening-Adjustierung und die Blockpermutation der modifizierten PC-Teststatistik die höchste Teststärke auf. Mit Werten zwischen 45 und 48 Prozent Güte liegen diese Verfahren hier etwa im Bereich der Güte der “univariaten” Analyse. Die weiteren multivariaten Teststatistiken folgen dann analog zum ersten Alternativszenario.

Simulation einer zufälligen räumlichen Abhängigkeitsstruktur

Die Güte der Teststatistiken und Adjustierungsverfahren für zeitliche Korrelation wird im Folgenden anhand der zweiten Alternativhypothese ($p_1 = 64$) für das Szenario untersucht, bei dem die räumliche Korrelationsstruktur jeweils zufällig simuliert wird. Bei doppelter Anzahl effektbehafteter Variablen steigen die Güteresultate für die Mehrzahl der Teststatistiken auf bis zu 100 Prozent an, so dass ein Vergleich der einzelnen Verfahren nicht aussagekräftig ist. Abbildung 6.19 stellt daher die Ergebnisse für dieses Simulationsszenario mit 64 Variablen unter der Alternative dar.

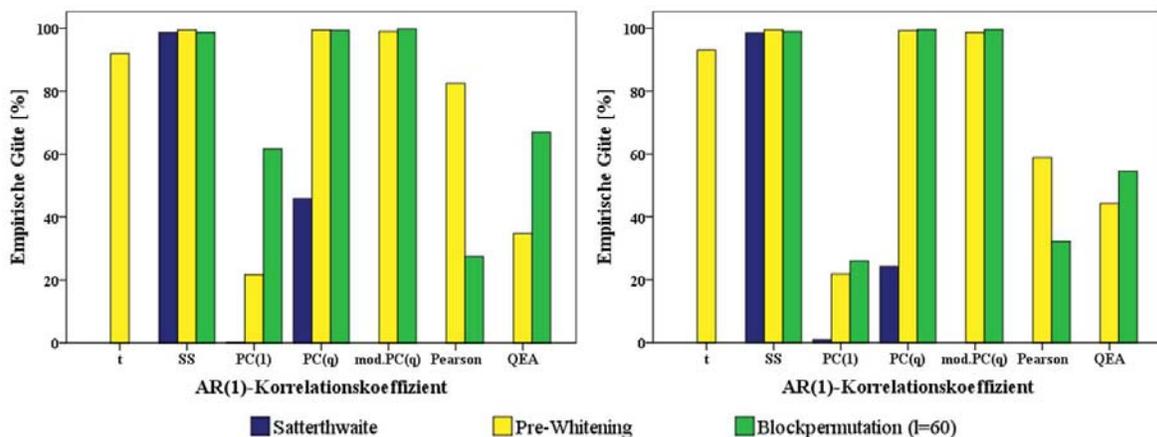


Abbildung 6.19.: Empirische Güte der verschiedenen Adjustierungsverfahren und Tests bei simulierten fMRT-Daten mit zufälliger räumlicher Korrelation; links Blockdesign, rechts ereignisbezogenes Design ($n = 250$, $p = 1024$, 64 Variablen unter der Alternative, $\rho = 0,4$, 299 Permutationen bzw. Rotationen, 2500 Wiederholungen)

Wie bereits bei den bisherigen Szenarien zur Gütesimulation erzielen die q -dimensionalen Principal Component Teststatistiken unter Pre-Whitening-Adjustierung eine sehr hohe Güte. Der klassische q -dimensionale PC-Test erreicht 99,4 Prozent Güte beim Blockdesign und 99,2 Prozent beim ereignisbezogenen Design; der modifizierte Test entsprechend 99 beziehungsweise 98,7 Prozent Güte. Die Dimension der Scores ist dabei mit einem

durchschnittlichen $q = 63$ etwa in Höhe der Variablenanzahl unter der Alternative, da es keine konkrete räumliche Struktur gibt, die Grundlage einer kleineren datengetriebenen Dimensionswahl bildet. Das erklärt auch, warum die Güte des eindimensionalen PC-Scores hier verhältnismäßig schlecht abschneidet, da innerhalb dieses einzelnen Scores nicht genug Informationen gebündelt werden können.

Der Standardisierte Summentest hingegen weist für dieses Szenario mit über 98,5 Prozent Güte für beide fMRT-Designs und alle Adjustierungsverfahren eine hohe Güte auf. Das könnte darin begründet sein, dass die zufällige Struktur durch die standardisierte Summierung besser erfasst werden kann als die im sonstigen Szenario simulierten 16 Gebiete. Zudem sind die simulierten Effekte jeweils gleichgerichtet.

Darüber hinaus zeigt die nichtmetrische Adjustierung für zeitliche Korrelation für den q -dimensionalen PC-Test mit 99,3 Prozent beim Blockdesign und 99,6 Prozent beim ereignisbezogenen Design keinen Güteverlust gegenüber dem Pre-Whitening. Dies gilt ebenso für die modifizierte q -dimensionalen PC-Teststatistik, die bei Verwendung des Blockpermutationsverfahrens Gütergebnisse von 99,8 beziehungsweise 99,6 Prozent erreicht.

Die auf Hauptkomponenten basierenden stabilisierten multivariaten Testverfahren liegen folglich mit Ausnahme der eindimensionalen PC-Teststatistik über der Güte des “univariaten” Verfahrens, die beim Blockdesign 91,9 Prozent, beim ereignisbezogenen Design 93,1 Prozent Güte erreicht. Anschließend folgt für beide fMRT-Designs die Güte des Pearson’schen Korrelationskoeffizienten unter Pre-Whitening sowie der Abstandsstatistik unter Verwendung der Blockpermutationsverfahrens.

Eine zufällige – und somit im übertragenen Sinn einheitliche, da unstrukturierte – räumliche Abhängigkeit, zudem bei gleichgerichteten Effekten, begünstigt demnach in diesem Szenario die Verwendung des Standardisierten Summentests. Der Test erzielt in dieser Simulation eine dem q -dimensionalen PC-Score vergleichbare Teststärke. Der eindimensionale PC-Score hingegen ist für dieses Szenario nicht aussagekräftig genug und die abstandsbasieren Maße können auch hier keinen Vorteil gegenüber dem “univariaten” Verfahren erzielen.

Durch die Simulationen in diesem Kapitel wird ersichtlich, dass die Verwendung multivariater Methoden – speziell der auf Hauptkomponenten basierenden stabilisierten multivariaten Verfahren – durchaus Gütevorteile bei der Analyse von fMRT-Daten erreichen kann. Generell erweist sich das klassische wie auch modifizierte q -dimensionale Principal Component Testverfahren dabei als gütestärkste Analyse, wohingegen die Ähnlichkeits- und Distanzmaße weitgehend unterlegen zu sein scheinen.

Ob zur Adjustierung der Verfahren an die zeitliche Korrelation der Daten tatsächlich die Pre-Whitening-Methode der in dieser Arbeit vorgestellten nichtmetrischen Alternative überlegen ist, bleibt in Abhängigkeit einer korrekten Korrelationsannahme jedoch fraglich und sollte daher anhand realer Datensätze überprüft werden.

7. Anwendung der Verfahren auf reale fMRT-Daten

In den bisherigen Kapiteln werden verschiedene multivariate Teststatistiken wie auch Adaptionsstrategien für eine zeitliche Korrelation zwischen den Stichprobenvektoren anhand simulierter Daten auf ihre Anwendbarkeit auf funktionelle MRT-Messungen untersucht. Die dabei simulierten zeitlichen wie räumlichen Strukturen sind zwar realen Daten nachempfunden, spiegeln diese aber nur idealisiert wider. Daher werden die Verfahren in diesem Kapitel nun auf reale Probandendaten angewendet und diesbezüglich evaluiert. Im Folgenden wird das Studiendesign vorgestellt und die Umsetzung der multivariaten Datenanalyse beschrieben. Dem schließt sich ein Vergleich der Studienergebnisse an. Die Resultate der multivariaten Analysen werden darüber hinaus in Bezug auf die univariaten Ergebnisse diskutiert.

7.1. Hintergrund der Studie zum Ultimatum Game

Die hier verwendeten Probandendaten entstanden im Rahmen einer fMRT-Studie zur Durchführung eines in der Ökonometrie bekannten sozialen Spiels: des sogenannten Ultimatum Games ([Hollmann, 2009](#); [Hollmann et al., 2011](#)). In diesem Spiel teilen sich zwei Spieler einen feststehenden Geldbetrag. Dabei ist ein Spieler der Angebotgeber, der dem anderen einen Teil des Betrages anbietet. Der Angebotnehmer entscheidet daraufhin, ob er das Angebot annimmt oder ablehnt. Bei Annahme wird der jeweilige Teilbetrag an die beiden Spieler ausgezahlt. Bei Ablehnung des Angebots erhält keiner der beiden Spieler einen Anteil.

Rein theoretisch müsste jeder Betrag vom Angebotnehmer angenommen werden, da jeder noch so kleine Gewinn keinem Gewinn vorzuziehen wäre. In der Realität zeigt sich jedoch, dass “unfaire” Angebote von maximal 20 Prozent des Geldbetrages in mehr als der Hälfte der Versuche abgelehnt werden. Somit ist die einzelne Entscheidung des Angebotnehmers nicht rein rational, sondern von Gefühlen und Erwartungen beeinflusst. [Hollmann et al. \(2011\)](#) untersuchten daher in einem Echtzeit-fMRT-Szenario, ob durch Nachweis negativer Emotionen beim Angebotnehmer vorhergesagt werden kann, ob er das Angebot ablehnen wird. Die Vorhersage der Entscheidung erfolgte hier bereits bevor der Proband seine Entscheidung kommunizierte.

Insgesamt wurden zehn Probanden in der Rolle des Angebotnehmers in einem 3Tesla Magnetresonanztomographen untersucht (Siemens Medical Systems, Erlangen). Ihnen wurden zuvor jeweils zwei weitere Probanden als vermeintliche Angebotgeber präsen-

tiert, deren Angebote im Verlauf des Experiments in zufälliger Abfolge erfolgen würden. Dies diente jedoch allein der Illusion einer persönlichen Interaktion. Die tatsächlichen Angebote erfolgten anschließend computergestützt in einer zuvor definierten Reihenfolge. Sämtliche Teilnehmer waren männlich, um einen Geschlechtereffekt zu vermeiden. Ein Proband musste aus der Analyse ausgeschlossen werden, da er im Anschluss Zweifel äußerte, gegen einen menschlichen Gegner gespielt zu haben und daher gewinnorientiert sämtliche Angebote angenommen hatte.

Jedem Probanden im MRT wurden nacheinander insgesamt 60 Angebote unterbreitet, wobei die Gesamtsumme jeweils drei Euro entsprach. In zufälliger Reihenfolge ergaben sich für jeden Probanden sechs Angebote mit 50:50-Aufteilung, achtmal 65:35, zwölfmal 70:30, 21-mal 80:20 und 13 Angebote zu 90:10. In der späteren Analyse werden die Angebote bis 35 Prozent als fair, solche mit maximal 20 Prozent Anteil als unfair bezeichnet. Ein einzelner Durchgang dauerte insgesamt 22 Sekunden, wobei die Zeit zwischen zwei Scans ein und derselben Schicht zwei Sekunden betrug. Innerhalb eines Durchgangs wurden dem Probanden zunächst zwei Sekunden lang die zu teilende Gesamtsumme angezeigt. Anschließend erfolgte für zwölf Sekunden die Einblendung des Angebots. Dem Probanden blieben dann vier Sekunden zum Antworten. Die Antwort wurde durch Drücken einer von zwei Tasten gegeben. Welche der beiden Tasten im jeweiligen Durchgang das Ablehnen des Angebots widerspiegelte, wechselte zufällig und wurde dem Probanden innerhalb dieser vier Sekunden angezeigt. Dies sollte verhindern, dass sich der Proband bereits während der Einblendung des Angebots darauf vorbereitete, die entsprechende Taste zu drücken, da dies unerwünschte Hirnaktivierungen während der Entscheidungsphase zur Folge gehabt hätte. Abschließend wurde das Ergebnis, welchen Summe beide Spieler erhalten, für vier Sekunden angezeigt.

Die lange Zeitspanne von zwölf Sekunden bis zur Kommunikation der Entscheidung war nötig, um währenddessen eine Echtzeitanalyse durchzuführen, also noch während der Einblendung des Angebots die Entscheidung des Probanden anhand seiner Hirnaktivierung vorhersagen zu können.

In der ursprünglichen Analyse von [Hollmann et al. \(2011\)](#) erfolgte die Echtzeit-Auswertung nicht wie standardmäßig für fMRT-Daten über die Analyse eines linearen Modells, sondern anhand einer Pattern-Klassifikation. Die Klassifikation der Probandenentscheidung wurde über eine sogenannte Relevance Vector Machine (RVM) realisiert. Dieser Klassifikator wurde zuvor anhand von Aktivierungsdaten zweier Probanden darauf trainiert, welche Aktivierungsmuster einer Ablehnung und welche einer Annahme des Angebots zuzuordnen waren.

Zur Klassifikation wurde das Aktivierungsschema dreier Gehirnareale genutzt, die alle unter anderem mit emotionalen Reaktionen assoziiert sind: die anteriore Insula, der laterale präfrontale Kortex und der visuelle Kortex. Die Aktivierungsmuster während der Entscheidungsphase wurden als Vektor zusammengefasst und schließlich der Entscheidung Ablehnung oder Annahme zugeordnet. Die Lage der drei Areale ist schematisch in [Abbildung 7.1](#) dargestellt. Die genaue Lokalisation der Areale ist aber für jeden Probanden geringfügig unterschiedlich und musste daher in Voruntersuchungen bestimmt werden.

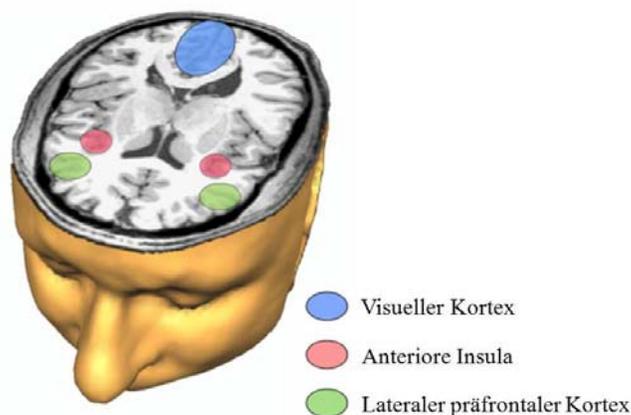


Abbildung 7.1.: Schematische Darstellung der Lokalisation der drei vordefinierten im Ultimatum Game referenzierten Gehirnareale (Hollmann, 2009)

Hintergrund der Verwendung der Klassifikationsanalyse bei dieser Umsetzung des Spiels in fMRT war einerseits, dass die Vorhersage in Echtzeit erfolgen musste und diese vektorbasierte Klassifikation schnell durchzuführen war. Dabei zeigte sich eine vom Rateniveau signifikant abweichende Klassifikationsgüte von durchschnittlich knapp 70 Prozent. Andererseits führten Offline-Analysen zu der Erkenntnis, dass die Aktivierungen für faire und unfaire Angebote anhand einer standardmäßigen voxelweisen Analyse kaum voneinander zu unterscheiden und damit ihre Aktivierungsdifferenzen kaum nachzuweisen waren.

Daher soll in diesem Kapitel untersucht werden, ob diese schwachen Aktivierungsunterschiede durch Verwendung der multivariaten Testverfahren besser nachweisbar sind.

7.2. Beschreibung der Daten

Die Daten der neun Probanden werden mit den in dieser Arbeit beschriebenen multivariaten Verfahren analysiert. Ziel ist es festzustellen, ob die Aktivierungen beim Vergleich von unfairen und fairen Angeboten überhaupt nachweisbar sind und wie detailliert die entsprechende Gehirnregion dabei referenziert werden muss.

Für jeden Probanden liegen $n = 660$ Messungen über insgesamt $53 \times 63 \times 46$ Voxel vor. Die Daten sind bewegungskorrigiert und nach MNI normalisiert (vgl. Abschnitt 3.1). Der interessierende Kontrast ergibt sich aus der Differenz der Aktivierung bei unfairen Angeboten (90:10 und 80:20) zu der bei fairen Angeboten (65:35 und 50:50). Beispielgebend wird die anteriore Insula rechts untersucht. Für den interessierenden Kontrast müssten dort Aktivierungen erkennbar sein, da diese Hirnregion mit negativen Emotionen assoziiert ist und somit die Aktivierung bei unfairen Angeboten stärker sein sollte als bei fairen Angeboten. Zur Lokalisation dieses Hirnareals wurde aus der univariaten Offline-Analyse mit dem Standardprogramm SPM (SPM, 2009) das lokale Aktivierungsmaximum der

Region bestimmt. Dabei ist zu beachten, dass in SPM standardmäßig einseitige Tests an räumlich geglätteten Daten durchgeführt werden. Das Aktivierungsmaximum muss daher nicht mit den hier durchgeführten Tests identisch sein, sondern dient nur der Lokalisation der Region. Für jeden Probanden wird an der individuell ausgewählten Stelle eine Region von $4 \times 4 \times 3$, also $p = 48$ Voxeln, bestimmt. Die anteriore Insula ist damit relativ genau referenziert. Weiterhin wird eine $p = 900$ Voxel ($10 \times 10 \times 9$) umfassende Hirnregion untersucht. Diese enthält die anteriore Insula, umfasst aber zusätzlich eine Vielzahl weiterer Voxel.

Die Designmatrix ist für jeden Probanden individuell und wird direkt aus dem Standardprogramm SPM übernommen. Sie besteht aus insgesamt zwölf Regressoren. Die ersten fünf spiegeln jeweils die einzelnen Angebote wieder – zunächst die fairen 50:50 und 65:35, dann 70:30, dann die unfairen 80:20 und 90:10. Anschließend folgen sechs Spalten zur Bewegungskorrektur in Form von Rotation und Translation für jede Raumrichtung (vgl. Abschnitt 3.1). Der letzte Regressor entspricht schließlich dem Gesamtmittel. Der interessierende Kontrast für die Aktivierungsdifferenz der unfairen und fairen Angebote ergibt sich somit aus

$$\mathbf{c}' = \begin{pmatrix} -1 & -1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

7.3. Ergebnisse der multivariaten Auswertung

Sämtliche multivariaten Teststatistiken unter Adjustierung für zeitliche Korrelation werden für jeden der neun Probanden sowohl für die individuell referenzierte rechte anteriore Insula wie auch für die sie umfassende größere Region durchgeführt. Ein Vergleich zur univariaten Analyse erfolgt über t -Tests unter Pre-Whitening-Adjustierung.

Auch hier wird jeweils eine AR(1)-Struktur für die zeitliche Korrelation der Messungen angenommen und der Korrelationskoeffizient aus den Daten geschätzt. Für die untersuchten fMRT-Daten liegt diese Korrelation bei den verschiedenen Probanden zwischen 0,121 und 0,309 und somit in einem Bereich, für den alle im Weiteren verwendeten multivariaten Teststatistiken unter Adjustierung der zeitlichen Korrelation das nominale Testniveau approximativ einhalten (vgl. Abschnitt 6.2.1).

Bei Verwendung des Blockpermutationsverfahrens wird jeweils die Blocklänge $l = 50$ unter Einbeziehung der zufälligen Verschiebung genutzt. Die Bestimmung der Score-dimension für die q -dimensionalen PC-Teststatistiken erfolgt über das inverse Jolliffe-Kriterium (2.22).

Die Tabelle in Abbildung 7.2 enthält die multivariaten p -Werte für alle Probanden sowie Testverfahren bei Untersuchung der individuellen rechten anterioren Insula. Für den univariaten Vergleich ist der minimale Bonferroni-korrigierte p -Wert der 48 t -Tests unter Pre-Whitening-Adjustierung angegeben. Die bei Verwendung einer Irrtumswahrscheinlichkeit von $\alpha = 0,05$ signifikanten Testergebnisse sind grün hinterlegt. Hinweise auf tendenzielle Unterschiede ($0,05 < p \leq 0,1$) sind gelb hervorgehoben.

Generell können die Ähnlichkeits- und Distanzmaße für keinen der neun Probanden die

Teststatistik	Adjustierung	Proband								
		1	2	3	4	5	6	7	8	9
t (min)	Pre-Whitening	1,0000	0,0010	0,0821	0,0019	0,0720	1,0000	0,2534	0,3816	0,0158
Standardisierter Summentest	Satterthwaite	0,5296	0,0019	0,1423	0,0018	0,0034	0,3486	0,0362	0,6419	0,0578
	Pre-Whitening	0,4660	0,0012	0,0719	0,0019	0,0011	0,4887	0,0258	0,5283	0,0246
	Blockpermutation	0,4470	0,0050	0,1010	0,0030	0,0020	0,3040	0,0560	0,6330	0,0550
	skalenabh. PC-Tests									
eindimensional	Satterthwaite	0,3046	0,0003	0,5605	0,0007	0,0111	0,0859	0,0795	0,4039	0,0412
	Pre-Whitening	0,2226	<0,0001	0,3511	0,0005	0,0026	0,1350	0,0836	0,2989	0,8313
	Blockpermutation	0,0980	0,0040	0,5740	0,0020	0,0480	0,0420	0,0660	0,4070	0,0580
	q -dimensional	0,8070	0,0149	0,0353	0,0068	0,0991	0,5903	0,5551	0,2232	0,1135
modifiziert	Pre-Whitening	0,7419	0,0081	0,0125	0,0055	0,0866	0,5855	0,5151	0,1964	0,0706
	Blockpermutation	0,6240	0,0100	0,0160	0,0060	0,0800	0,2850	0,4930	0,1300	0,0310
	Pre-Whitening	0,2460	0,0010	0,2510	0,0010	0,0030	0,2120	0,1000	0,3410	0,0440
	Blockpermutation	0,1330	0,0040	0,4160	0,0030	0,0450	0,0560	0,0760	0,4240	0,0530
skaleninvar. PC-Tests										
	eindimensional	0,3680	0,0005	0,1779	0,0007	0,0090	0,1479	0,0485	0,5534	0,0079
	Pre-Whitening	0,2608	0,0001	0,0985	0,0007	0,0023	0,1769	0,0427	0,4209	0,0018
	Blockpermutation	0,1780	0,0040	0,1500	0,0020	0,0270	0,0810	0,0550	0,5440	0,0120
q -dimensional	Satterthwaite	0,9647	0,0897	0,0789	0,0198	0,1234	0,5173	0,3567	0,5478	0,0186
	Pre-Whitening	0,9576	0,0124	0,0125	0,0312	0,0732	0,6437	0,3166	0,2474	0,0121
	Blockpermutation	0,9010	0,0420	0,0190	0,0260	0,0760	0,1890	0,2680	0,3730	0,0050
	modifiziert	0,3750	0,0010	0,0940	0,0010	0,0040	0,3350	0,0800	0,4740	0,0020
Pearson'scher Korrelationskoeff. quadr. euklidischer Abstand	Blockpermutation	0,3590	0,0040	0,1510	0,0020	0,0270	0,0970	0,0890	0,5590	0,0130
	Pre-Whitening	0,1740	0,1580	0,6190	0,7310	0,7420	0,7720	0,3200	0,9420	0,5420
	Blockpermutation	0,1550	0,1620	0,5630	0,7400	0,6620	0,7510	0,2850	0,8510	0,4810
	Pre-Whitening	0,4070	0,5280	0,7590	0,8360	0,7900	0,8310	0,6760	0,8240	0,5550
Abstand	Blockpermutation	0,1650	0,1320	0,5130	0,5960	0,7750	0,8100	0,4390	0,7490	0,1660

Abbildung 7.2.: Ergebnisse der Tests auf Aktivierungsunterschiede von unfairen und fairen Angeboten innerhalb der rechten anterioren Insula für verschiedene Teststatistiken und Adjustierungsverfahren ($p = 48$, $l = 50$, 999 Permutationen bzw. Rotationen)

Aktivierungsdifferenz von unfairen und fairen Angeboten nachweisen. Dies gilt sowohl in der individuell referenzierten anterioren Insula wie auch für das größere, sie umfassende Gebiet (vgl. Tabelle in Abbildung 7.3).

Bei den Probanden 2 und 4 sind die Aktivierungen in der rechten anterioren Insula stark genug, um bereits univariat unter Bonferroni-Korrektur jeweils fünf aktivierte Voxel nachzuweisen. Analog dazu weisen auch die multivariaten Teststatistiken größtenteils eine Aktivierungsdifferenz für die unfairen und fairen Angebote nach.

Für Proband 2 liegen die Ergebnisse der Standardisierten Summenteststatistik wie auch der eindimensionalen PC-Teststatistiken für sämtliche Adjustierungsverfahren für zeitliche Korrelation maximal bei $p = 0,005$. Die Ergebnisse der q -dimensionalen PC-Teststatistiken liegen mit Werten zwischen $p = 0,0081$ und $p = 0,0897$ ein wenig höher, weisen jedoch mit einer Ausnahme auch die signifikante Aktivierung im referenzierten Gebiet nach. Die Dimension der Scores ergibt sich dabei für die skalenabhängige Version aus $q = 6$ bei Satterthwaite-Approximation beziehungsweise $q = 8$ bei Pre-Whitening. Für die skaleninvariante Testversion ist jeweils $q = 9$. Bei Bestimmung des p -Wertes für die modifizierten PC-Teststatistiken über je 999 Rotationen entsteht keine extremere Konstellation gegen die Nullhypothese, so dass sich der hier kleinstmögliche p -Wert von $p = 0,001$ ergibt. Das Blockpermutationsverfahren erzielt geringfügig höhere p -Werte.

Proband 4 weist im Vergleich dazu nur minimale Unterschiede in den Ergebnissen auf. Hier können alle auf Hauptkomponenten basierenden stabilisierten multivariaten Testverfahren eine Aktivierung in der anterioren Insula nachweisen. Die Dimension der PC-Scores liegt dabei jeweils bei $q = 8$ beziehungsweise für die Pre-Whitening-Version der skaleninvarianten Teststatistik bei $q = 9$.

Generell sind bei diesen Probanden keine systematischen Unterschiede zwischen den metrischen Adjustierungsverfahren für zeitlichen Korrelation und dem Blockpermutationsverfahren mit zufälliger Verschiebung festzustellen.

Auch bei Proband 9 ist die Aktivierungsdifferenz in der rechten anterioren Insula hoch genug, um univariat zumindest ein einzelnes unter Bonferroni-Korrektur signifikantes Voxel nachzuweisen. Die Ergebnisse der multivariaten Verfahren sind dabei allerdings weniger eindeutig als bei den Probanden 2 und 4. Der Standardisierte Summentest weist die Aktivierung nur unter Pre-Whitening-Adjustierung signifikant nach, während bei Satterthwaite-Approximation und dem Blockpermutationsverfahren mit $p = 0,0578$ beziehungsweise $p = 0,0550$ nur eine Tendenz erkennbar ist. Der eindimensionale skalenabhängige PC-Test hingegen kann die Aktivierung unter Pre-Whitening gar nicht, bei Verwendung der blockweisen Permutation tendenziell und unter Satterthwaite signifikant nachweisen. In der skaleninvarianten Version gelingt der Nachweis der Aktivierungsdifferenz indes für alle Adjustierungen an die zeitliche Korrelation. Dies gilt ebenfalls für den skaleninvarianten q -dimensionalen PC-Test, dessen Scoredimension jeweils bei $q = 8$ liegt. Der skalenabhängige PC-Test hingegen liefert nur anhand des Blockpermutationsverfahrens eine Signifikanz.

Für die Probanden 5 und 7 können innerhalb der individuell referenzierten anterioren In-

sula univariat nur unadjustiert einzelne Voxel als signifikant bestimmt werden, Proband 5 weist eine tendenzielle Aktivierung in einem Voxel auf. Die multivariaten Verfahren können indes die Aktivierungsdifferenz der unfairen und fairen Angebote vielfach nachweisen. Sämtliche Standardisierte Summenteststatistiken wie auch die eindimensionalen PC-Tests liefern bei Proband 5 für alle drei zeitlichen Adjustierungsverfahren signifikante Ergebnisse. Tendenziell ist dies genauso bei Proband 7 erkennbar, wobei die entsprechenden Ergebnisse teilweise leicht über dem Testniveau von fünf Prozent liegen.

Univariat lassen sich für Proband 3 unadjustiert in fünf Voxeln Aktivierungsdifferenzen zwischen unfairen und fairen Angeboten nachweisen. Beim in der fMRT-Analyse üblichen unadjustierten Testniveau von 0,1 Prozent oder bei Bonferroni-Korrektur können diese Ergebnisse jedoch nicht verifiziert werden. Hier bleibt lediglich eine Tendenz, während einige multivariate Verfahren die Aktivierung signifikant nachweisen können. In diesem Fall gilt dies vor allem für den q -dimensionalen PC-Test ($6 \leq q \leq 8$). Die skalenabhängige Version liefert bei allen Adjustierungsverfahren für zeitliche Korrelation einen signifikanten Nachweis der Aktivierungsdifferenz. Die skaleninvariante Version erreicht unter Pre-Whitening sowie der blockweisen Permutation ein signifikantes Ergebnis. Das Pre-Whitening scheint bei diesem Probanden am sensitivsten zu sein.

Für die Probanden 1 und 6 können univariat keine aktivierten Voxel in der rechten anterioren Insula festgestellt werden. Der kleinste unadjustierte p -Wert der t -Tests unter Pre-Whitening liegt für beide Probanden über 0,06. Bei Proband 6 ergibt sich bei Nutzung der eindimensionalen skalenabhängigen Principal Component Teststatistik eine Signifikanz mit $p = 0,042$ bei Verwendung der Blockpermutation mit Random Shift. Für die gleiche Teststatistik kann unter Satterthwaite-Approximation zumindest ein tendenzielles Ergebnis von $p = 0,0859$ erzielt werden. Die entsprechende skaleninvariante Teststatistik erreicht beim blockweisen Permutationsverfahren ebenfalls eine Tendenz. Für Proband 1 ist mit $p = 0,098$ für den eindimensionalen skalenabhängigen PC-Test unter Verwendung der nichtmetrischen Adjustierung für zeitliche Korrelation eine tendenzielle Aktivierung zu finden. Alle anderen Teststatistiken finden keine Aktivierungsdifferenz der unfairen und fairen Angebote für diese beiden Probanden. Die Dimension der PC-Scores liegt hier aber vergleichbar mit allen anderen Probanden auch zwischen $q = 6$ und $q = 9$.

Schließlich erfolgt für Proband 8 bei Test der Differenz von unfairen und fairen Angeboten weder univariat unter multipler Adjustierung oder verringertem unkorrigierten Testniveau noch multivariat ein Nachweis der Aktivierung in der rechten anterioren Insula. Lediglich unkorrigiert liegen drei der univariaten p -Werte unterhalb des Testniveaus von 0,05. Den kleinsten multivariaten p -Wert der hier durchgeführten Analysen liefert mit $p = 0,13$ der q -dimensionale skalenabhängige PC-Test bei Verwendung der blockweisen Permutation mit zufälliger Verschiebung.

Zusammenfassend kann für die probandenspezifisch referenzierte rechte anteriore Insula mit jeweils $p = 48$ Voxeln der Nachweis der Aktivierungsdifferenz von unfairen und fairen Angeboten während des Ultimatum Games durch Verwendung der multivariaten

Verfahren zum Teil verbessert werden (vgl. Tabelle in Abbildung 7.2). Vor allem die auf Hauptkomponenten basierenden stabilisierten multivariaten Verfahren zeigen in der Anwendung auf diese kleine homogene Hirnregion Gütevorteile, während die Ähnlichkeits- und Distanzmaße keine Aktivierungen nachweisen können.

Dort, wo durch univariate Analysen bereits einzelne Voxel als signifikant bestimmt werden können, finden sämtliche auf Hauptkomponenten basierende stabilisierte multivariate Verfahren und Adjustierungen für zeitliche Korrelation diese Aktivierung ebenfalls. Zwar gibt es auch Probanden, für die die geringen Aktivierungsunterschiede auch multivariat gar nicht oder kaum nachgewiesen werden können. Bei den übrigen Probanden, bei denen die univariate Analyse unter Einhaltung des multiplen Testniveaus durch Bonferroni-Korrektur keinen Aktivierungsnachweis erzielt, ist dieser Nachweis vor allem mit dem Standardisierten Summentest wie auch dem eindimensionalen PC-Teststatistiken und vereinzelt über die höherdimensionalen Scores möglich. Gerade für die q -dimensionalen Scores scheint die modifizierte Version der Teststatistik hier vorteilhaft zu sein.

Die einzelnen Adjustierungsmethoden für zeitliche Korrelation weisen dabei keine systematischen Unterschiede zueinander auf.

Die Tabelle in Abbildung 7.3 enthält die entsprechenden Ergebnisse für die neun Probanden bei Untersuchung der größer gewählten Hirnregion, die neben der rechten anterioren Insula eine Vielzahl umliegender Voxel umfasst.

Selbst für diese mit 900 Voxeln vergleichsweise große und somit heterogene Region reichen die starken Aktivierungen bei den Probanden 2 und 4 aus, univariat unter Bonferroni-Adjustierung acht beziehungsweise sechs signifikant aktivierte Voxel nachzuweisen. Multivariat werden diese Aktivierungen hauptsächlich durch die q -dimensionalen Principal Component Teststatistiken nachgewiesen.

Für Proband 2 ist das Ergebnis der q -dimensionalen skalenabhängigen PC-Teststatistik bei allen drei Adjustierungsverfahren für zeitliche Korrelation signifikant. Die Scoredimension liegt dabei bei $q = 51$ (Satterthwaite) beziehungsweise $q = 55$ (Pre-Whitening). Das skaleninvariante Pendant, dessen Dimension der Scores auf $q = 92$ beziehungsweise $q = 99$ steigt, erreicht nur bei Verwendung des Blockpermutationsverfahrens sowie als modifizierte Testversion unter Pre-Whitening oder blockweiser Permutation ein signifikantes Ergebnis. Trotz der Größe und damit einhergehenden Heterogenität der Region liefert auch der Standardisierte Summentest für diesen Probanden in allen drei Adjustierungsversionen einen signifikanten Nachweis der Aktivierungsdifferenz von unfairen und fairen Angeboten. Analog dazu zeigt die skaleninvariante eindimensionale PC-Teststatistik unter Satterthwaite-Approximation sowie bei Verwendung des Blockpermutationsverfahrens eine Tendenz auf. Dies könnte darin begründet sein, dass die Aktivierungen des Probanden sehr stark und einheitlich sind und die aktivierten Voxel der anterioren Insula diese Hirnregion dominieren. Darauf deutet auch die mittlere Korrelation der Voxel von $\bar{\rho}_r = 0,239$ hin. Der mittlere Korrelationskoeffizient $\bar{\rho}_r$ wird dabei aus der $p \times p$ -dimensionalen geschätzten Korrelationsmatrix der Voxel ohne Einbeziehung der Hauptdiagonale berechnet. Er soll hier nur einen Hinweis auf die durchschnittliche

Teststatistik	Adjustierung	Proband									
		1	2	3	4	5	6	7	8	9	
t (min)	Pre-Whitening	1,0000	0,0009	1,0000	0,0036	0,2538	1,0000	0,3591	0,8991	0,1413	
Standardisierter Summentest	Satterthwaite	0,9953	0,0184	0,8739	0,2902	0,2365	0,1846	0,2117	0,0419	0,5456	
	Pre-Whitening	0,9908	0,0131	0,7735	0,3060	0,1661	0,1268	0,1060	0,0328	0,5344	
	Blockpermutation	0,9940	0,0420	0,8540	0,3310	0,1910	0,1480	0,2500	0,0630	0,5640	
	skalenabh. PC-Tests										
eindimensional	Satterthwaite	0,7878	0,3887	0,9016	0,1273	0,4457	0,8764	0,2093	0,2116	0,5253	
	Pre-Whitening	0,5146	0,3586	0,7660	0,1795	0,3373	0,7967	0,1247	0,2248	0,5561	
	Blockpermutation	0,6300	0,3420	0,8970	0,1660	0,4010	0,8670	0,2280	0,2730	0,4670	
	Satterthwaite	0,9860	0,0057	0,2223	0,0001	0,1007	0,8694	0,7105	0,1710	0,0134	
q -dimensional	Pre-Whitening	0,9253	0,0006	0,0445	<0,0001	0,0224	0,8224	0,4747	0,2071	0,0141	
	Blockpermutation	0,6760	0,0040	0,0210	0,0020	0,0350	0,0760	0,2800	0,0200	0,0120	
	Pre-Whitening	0,8070	0,1960	0,7700	0,0350	0,6230	0,2230	0,3450	0,1400	0,5490	
	Blockpermutation	0,7880	0,1920	0,7850	0,0580	0,7600	0,2480	0,4020	0,1960	0,4540	
skalenin. PC-Tests											
	eindimensional	Satterthwaite	0,8581	0,0675	0,8664	0,2680	0,3213	0,6396	0,2844	0,0830	0,2577
	Pre-Whitening	0,8001	0,4916	0,7505	0,3444	0,2641	0,4087	0,1482	0,0841	0,2687	
	Blockpermutation	0,8180	0,0980	0,8520	0,3180	0,2820	0,5960	0,3230	0,1160	0,2700	
q -dimensional	Satterthwaite	0,9993	0,4611	0,6662	0,0022	0,1527	0,9933	0,8705	0,1732	0,0080	
	Pre-Whitening	0,9993	0,1045	0,4424	0,0006	0,1111	0,9872	0,6027	0,1964	0,0069	
	Blockpermutation	0,7220	0,0070	0,0360	0,0010	0,0090	0,1470	0,2240	0,0030	0,0030	
	Pre-Whitening	0,7710	0,0150	0,7910	0,0460	0,4730	0,1650	0,3490	0,0530	0,3460	
modifiziert	Blockpermutation	0,7110	0,0480	0,7700	0,0720	0,5780	0,1530	0,4620	0,1010	0,3290	
	Pre-Whitening	0,2570	0,5630	0,7210	0,6850	0,9980	0,6690	0,2470	0,5510	0,2930	
	Blockpermutation	0,2300	0,6110	0,6180	0,7030	0,9560	0,6810	0,2280	0,5180	0,2700	
	Pre-Whitening	0,5780	0,6450	0,7920	0,8080	0,9520	0,7120	0,5930	0,6700	0,5480	
quadr. euklidischer Abstand	Blockpermutation	0,2770	0,4720	0,5520	0,5270	0,8370	0,6020	0,2590	0,4750	0,2160	

Abbildung 7.3.: Ergebnisse der Tests auf Aktivierungsunterschiede von unfairen und fairen Angeboten innerhalb der größeren Region inklusive der rechten anterioren Insula für verschiedene Teststatistiken und Adjustierungsverfahren ($p = 900$, $l = 50$, 999 Permutationen bzw. Rotationen)

Stärke der räumlichen Korrelation geben, wobei die Richtung der Korrelation wie auch die Lage der Variablen zueinander unberücksichtigt bleiben. Neben Proband 2 weisen die Probanden 6 und 8 mit $\bar{\rho}_r = 0,291$ beziehungsweise $\bar{\rho}_r = 0,254$ relativ hohe mittlere räumliche Korrelationen auf, die im Bereich derer innerhalb der genau referenzierten anterioren Insula der neun Probanden $[0,182;0,430]$ liegen. Die restlichen Probanden haben dagegen in der größeren Region eine durchschnittliche räumliche Korrelation zwischen 0,067 und 0,141. Eben für diese drei Probanden mit höher korrelierten Voxeln weist der Standardisierte Summentest im Vergleich zu den übrigen Teststatistiken relativ kleine p -Werte auf.

Für Proband 4 wird die Aktivierung im Vergleich der unfairen und fairen Angebote indes allein durch die q -dimensionalen Teststatistiken nachgewiesen. Die Ergebnisse der skalenabhängigen Version, deren Scoredimension bei $q = 59$ (Satterthwaite) beziehungsweise $q = 61$ (Pre-Whitening) liegt, unterscheiden sich dabei nur minimal von denen der skaleninvarianten Version, deren Scores 99- beziehungsweise 100-dimensional sind. Auch die verschiedenen Adjustierungsverfahren für zeitliche Korrelation weichen nur unwesentlich voneinander ab. Lediglich die modifizierten PC-Teststatistiken weisen vergleichsweise höhere p -Werte auf. Keine der weiteren multivariaten Teststatistiken kann jedoch die Aktivierung innerhalb der großen Region nachweisen.

Für Proband 9, für den univariat weder bei unkorrigiertem Testniveau von 0,1 Prozent noch unter Bonferroni-Korrektur der Nachweis eines signifikanten Voxels möglich ist, zeigen lediglich die klassischen q -dimensionalen PC-Teststatistiken die Aktivierungsdifferenz an. Die Ergebnisse der skaleninvarianten Teststatistik, deren Scores die Dimension 67 (Satterthwaite) beziehungsweise 71 (Pre-Whitening) besitzen, sind dabei für alle Adjustierungsverfahren für zeitliche Korrelation ein bisschen kleiner als die der skalenabhängigen Version basierend auf 41- beziehungsweise 40-dimensionalen Scores. Unterschiede zwischen den Adjustierungsverfahren existieren dabei nicht.

Abweichungen zwischen den drei Adjustierungsverfahren sind hingegen für die Probanden 3 und 5 erkennbar. Auch hier wird die Differenz der Aktivierungsstärken von unfairen und fairen Angeboten in der großen Region nur von den klassischen q -dimensionalen PC-Teststatistiken nachgewiesen – von beiden Varianten jedoch ausschließlich bei Verwendung der blockweisen Permutation mit Random Shift. Bei Pre-Whitening-Korrektur kann nur durch die skalenabhängige q -dimensionale PC-Teststatistik ($q = 66$ bei Proband 3, $q = 55$ bei Proband 5) eine Aktivierungsdifferenz nachgewiesen werden.

Ähnlich verhält es sich bei Proband 8. Auch hier erzielen die klassischen q -dimensionalen PC-Teststatistiken bei Durchführung der blockweisen Permutation mit zufälliger Verschiebung einen Aktivierungsnachweis. Eine tendenzielle Aktivierung wird zudem durch die modifizierte q -dimensionale ($q = 111$) skaleninvariante PC-Teststatistik unter Pre-Whitening gefunden. Hinzu kommt bei diesem Probanden, dessen referenzierte große Hirnregion mit $\bar{\rho}_r = 0,254$ eine relativ starke und somit einheitliche räumliche Korrelation aufweist, dass die Standardisierte Summenteststatistik unter Satterthwaite-Approximation wie auch Pre-Whitening eine Signifikanz, bei Verwendung des Block-

permutationsverfahrens eine Tendenz aufzeigt. Weiterhin wird auch durch die eindimensionale skaleninvariante PC-Teststatistik sowohl unter Satterthwaite wie auch Pre-Whitening eine tendenzielle Aktivierung erkannt.

Für Proband 6 kann einzig die q -dimensionale skalenabhängige PC-Teststatistik bei Verwendung des Blockpermutationsverfahrens eine Tendenz des Aktivierungsunterschiedes von unfairen und fairen Angeboten innerhalb der großen Hirnregion aufzeigen.

Für die Probanden 1 und 7 wird innerhalb der Region mit 900 Voxeln keinerlei Aktivierungsdifferenz festgestellt.

Letztlich kann für die große untersuchte Hirnregion, die neben der anterioren Insula eine Vielzahl umliegender Voxel beinhaltet, der Nachweis einer unterschiedlichen Aktivierungsstärke von unfairen und fairen Angeboten durch die multivariaten Testverfahren größtenteils verbessert werden (vgl. Tabelle in Abbildung 7.3). Die Aktivierungen werden hierbei vor allem anhand der klassischen q -dimensionalen Principal Component Teststatistiken nachgewiesen, während die Modifizierung der Teststatistik keine Vorteile bringt. Die Scoredimensionen sind dabei aufgrund der Anzahl wie auch Heterogenität der untersuchten Voxel wesentlich höher als bei Analyse der individuell referenzierten Gebiete. Generell weisen auch hier die verschiedenen Adjustierungsverfahren für zeitliche Korrelation kaum Unterschiede zueinander auf. Lediglich für die Probanden, bei denen die Aktivierung innerhalb der untersuchten Region überhaupt nur tendenziell erkennbar ist, wird dies vor allem bei Verwendung der blockweisen Permutation mit zufälliger Verschiebung realisiert. Dies könnte am Vorteil des Verfahrens liegen, unabhängig von einer konkreten zeitlichen Korrelationsstrukturannahme zu sein.

Zusammenfassung der Ergebnisse:

Generell kann die Anwendung der multivariaten Methoden zur besseren Erkennung schwacher Aktivierungen führen. Dabei erweisen sich vor allem die auf Hauptkomponenten basierenden stabilisierten multivariaten Verfahren als hilfreich, während die Ähnlichkeits- und Distanzmaße für fMRT-Daten ungeeignet erscheinen. Dies gilt sowohl für kleine Hirnregionen, die relativ homogen aktiviert sind, wie auch für größere Areale, in denen sich eine kleine aktivierte Region befindet.

Entsprechend dieser Unterscheidung der Areale sind verschiedene multivariate Verfahren vorteilhaft: Speziell in gut referenzierten kleinen Gebieten weisen der Standardisierte Summentest wie auch die eindimensionalen Principal Component Teststatistiken Vorteile auf, wobei für die untersuchten Probanden keine eindeutige Präferenz für die skalenabhängige oder -invariante Testversion feststellbar ist. Zudem unterscheiden sich die verschiedenen Adjustierungsmethoden für zeitliche Korrelation für die im Beispiel untersuchten Daten nicht.

Bei der Analyse einer unpräzise gewählten Region ist vor allem anhand des q -dimensionalen PC-Tests eine Aktivierung besser erkennbar. Hierbei liefert insbesondere das Blockpermutationsverfahren mit zufälliger Verschiebung gute Ergebnisse, das ohne konkrete Annahmen einer zeitlichen Korrelationsstruktur auskommt.

8. Resümee

In den vorherigen Kapiteln wurde die Anwendung multivariater Verfahren auf funktionelle MRT-Daten zum Nachweis neuronaler Aktivität unter Adjustierung der zeitlichen Korrelation erfolgreich realisiert. Dies umfasste eine Adaption der in univariaten fMRT-Analysen etablierten Adjustierungsverfahren für die zeitliche Abhängigkeit der Stichprobenvektoren wie auch die Einführung einer nichtmetrischen Alternative für diese Verfahren. Es wurde gezeigt, dass das nominale Testniveau dabei für alle untersuchten Teststatistiken sowohl unter Satterthwaite-Approximation und Pre-Whitening wie auch bei Verwendung der hier vorgeschlagenen blockweisen Permutation mit zufälliger Verschiebung approximativ eingehalten wird. Wenige Ausnahmen traten für zeitliche Korrelationsstärken auf, die in realen Messungen kaum erreicht werden. Selbst dann ermöglichte eine Kopplung von Pre-Whitening mit dem Blockpermutationsverfahren das approximative Einhalten des nominalen Testniveaus.

Während die metrischen Adjustierungsverfahren für zeitliche Korrelation an eine konkrete Annahme der Abhängigkeitsstruktur, standardmäßig in Form eines autoregressiven Prozesses erster Ordnung, gebunden sind, ist dies für die in dieser Arbeit vorgeschlagene blockweise Permutationsstrategie nicht notwendig. Allein über die Annahme einer abnehmenden zeitlichen Korrelation für weiter voneinander entfernte Messungen wird durch eine geeignete Wahl der Blocklänge eine approximative Unabhängigkeit der Blöcke erzielt. Dabei gewährleistet die zufällige Verschiebung (Random Shift), dass auch bei relativ geringer Anzahl zu permutierender Blöcke genug unterschiedliche Kombinationen möglich sind, um die Permutationsverteilung der multivariaten Teststatistik ausreichend zu charakterisieren.

Durch Untersuchungen anhand simulierter fMRT-Daten wurde zudem dargestellt, dass vor allem die auf Hauptkomponenten basierenden stabilisierten multivariaten Verfahren Gütevorteile bei der Erkennung aktivierter Hirnareale gegenüber der univariaten, multiple adjustierten Analyse aufweisen, während die Verwendung von Ähnlichkeits- und Distanzmaßen sich als ungeeignet herausstellten.

Weiterhin verdeutlichte die Anwendung der Verfahren auf reale Datensätze, dass sich die Vorteile der einzelnen Teststatistiken in Abhängigkeit von Größe und Struktur der untersuchten Hirnregionen präzisieren: Gerade in kleinen, funktionell zusammenhängenden Hirnarealen erwies sich neben den auf eindimensionalen Scores basierenden Principal Component Teststatistiken der Standardisierte Summentest als vorteilhaft. Dies kann vor allem damit begründet werden, dass die untersuchten neuronalen Aktivierungen in funktionellen Gebieten gleichgerichtet und denkbar einheitlich sind. Dem gegenüber sind

in größeren Hirnarealen, die neben einer aktivierten Region eine Vielzahl umliegender Voxel enthalten können, die q -dimensionalen Principal Component Teststatistiken vorzuziehen, die sich hierbei als gütestärkste Verfahren zum Nachweis neuronaler Aktivität herausgestellt. Die Dimension der ihnen zugrunde liegenden Scores wurde anhand des inversen Jolliffe-Kriteriums in Abhängigkeit von den Daten selbst bestimmt. Die modifizierte Form dieser Teststatistik, welche die q -dimensionalen Scores in eine eindimensionale Teststatistik transformiert, schien dabei eher in kleineren, einheitlichen Gebieten vorteilhaft gegenüber der klassischen Variante zu sein.

Zwar zeigte sich in den Simulationen das Pre-Whitening zur Adjustierung der Testverfahren an die zeitliche Korrelation der Daten als gütestärkste Methode. In den Simulationsszenarien entsprach jedoch die angenommene zeitliche Korrelationsstruktur auch tatsächlich dem in den Daten simulierten autoregressiven Modell erster Ordnung. In den Analysen der realen fMRT-Daten manifestierte sich indes keine systematische Unterlegenheit des Blockpermutationsverfahrens gegenüber der metrischen Adjustierungsmethoden. Speziell in den großen, heterogenen Hirnarealen wurde anhand des in dieser Arbeit vorgestellten nichtmetrischen Verfahrens am häufigsten eine neuronale Aktivierung nachgewiesen.

Die vorliegende Arbeit verdeutlicht, dass sich die Anwendung stabilisierter multivariater Teststatistiken als vorteilhaft erweist, insbesondere um schwache Aktivierungen im Gehirn zu identifizieren. Je nach Größe und Struktur der untersuchten Regionen können die verschiedenen Realisierungen dieser Testverfahren optimal ausgenutzt werden. Neben den etablierten Anpassungsstrategien an die Korrelation der Stichprobenvektoren stellt sich die hier erarbeitete blockweise Permutation mit Random Shift als adäquate Alternative heraus, deren Vorteil in der unkonkreteren Annahme der Abhängigkeitsstruktur liegt. Zwar benötigt das Verfahren zur Erzeugung der Permutationsverteilung eine längere Rechenzeit als die einfache Schätzung eines Korrelationskoeffizienten bei Durchführung des Pre-Whitenings. Eine Optimierung dieser Korrelationsschätzung wäre jedoch ebenso rechenintensiv.

Folglich bieten die multivariaten Verfahren die Möglichkeit, selbst ohne exakte Referenzierung des Hirnareals schwache Aktivierungen nachzuweisen, die durch univariate, auf multiples Testen adjustierte Analysen nicht identifiziert werden. Dennoch ist die Anwendung multivariater Verfahren nur sinnvoll, wenn zumindest grob festgelegt werden kann, in welchem Hirnareal eine Aktivierung zu erwarten ist. Weitere Entwicklungen sollten deshalb die Suche nach geeigneten Clustern von Voxeln im Rahmen multipler Testprozeduren einbeziehen, wie sie – ohne die Problematik korrelierter Stichprobenelemente – in den Arbeiten von [Läuter *et al.* \(2005\)](#) oder [Meinshausen \(2008\)](#) realisiert sind.

Literaturverzeichnis

- ADOLF, D, BAECKE, S, KAHLE, W, BERNARDING, J, & KROPF, S. 2011. Applying multivariate techniques to high-dimensional temporally correlated fMRI data. *Journal of Statistical Planning and Inference*, **141**, 3760–3770.
- AHMAD, MR. 2008. *Analysis of High Dimensional Repeated Measures Designs: The One- and Two-Sample Test Statistics*. Dissertation, Georg-August-Universität zu Göttingen.
- AHRENS, H, & LÄUTER, J. 1981. *Multivariate Varianzanalyse*. 2 edn. Berlin: Akademie-Verlag.
- ANDERSON, MJ. 2001. A new method for non-parametric multivariate analysis of variance. *Australian Ecology*, **26**, 32–46.
- ANDERSON, TW. 1984. *An Introduction to Multivariate Statistical Analysis*. 2 edn. New York: John Wiley & Sons.
- BAECKE, S, LÜTZKENDORF, R, TEMPELMANN, C, MÜLLER, C, ADOLF, D, SCHOLZ, M, & BERNARDING, J. 2009. Event-related functional magnetic resonance imaging (efMRI) of depth-by-disparity perception: additional evidence for right-hemispheric lateralization. *Experimental Brain Research*, **196**, 453–458.
- BASSO, D, PESARIN, F, SALMASO, L, & SOLARI, A. 2009. *Permutation Tests for Stochastic Ordering and ANOVA: Theory and Applications with R*. Berlin: Springer.
- BENNETT, CM, BAIRD, AA, MILLER, MB, & WOLFORD, GL. 2009. Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Argument For Proper Multiple Comparisons Correction. *Journal of Serendipitous and Unexpected Results*, **1**(1), 1–5.
- BERNARDO, JM. 1996. The Concept of Exchangeability and its Applications. *Far East Journal of Mathematical Sciences*, 111–121. Special Volume, Part II.
- BOYNTON, GM, ENGEL, SA, GLOVER, GH, & HEEGER, DJ. 1996. Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1. *The Journal of Neuroscience*, **16**(13), 4207–4221.
- BRAININNOVATION. 2011. *Brain Voyager QX 2.9*. <http://www.brainvoyager.com/products/brainvoyagerqx.html>, zuletzt geprüft am 25.11.2011: Brain Innovation B.V.

- CHUMBLEY, J, WORSLEY, KJ, FLANDIN, G, & FRISTON, KJ. 2010. Topological FDR for neuroimaging. *NeuroImage*, **49**(4), 3057–3064.
- DAVIS, CS. 2002. *Statistical methods for the analysis of repeated measurements*. New York: Springer.
- DING, GC, SMALLA, K, HEUER, H, & KROPF, S. 2012. A new proposal for a principal component based test for high-dimensional data applied to the analysis of PhyloChip data. *Biometrical Journal*, **54**(1), 94–107.
- DÖSSEL, O. 2000. *Bildgebende Verfahren in der Medizin: von der Technik zur medizinischen Anwendung*. Berlin Heidelberg: Springer.
- EFRON, B. 1977 (July). *Bootstrap Methods: Another Look at the Jackknife*. Technical report No. 32.
- FAHRMEIR, L, HAMERLE, A, & TUTZ, G. 1996. *Multivariate statistische Verfahren*. 2 edn. Berlin New York: de Gruyter.
- FAHRMEIR, L, KÜNSTLER, R, PIGEOT, I, & TUTZ, G. 2003. *Statistik*. 4 edn. Berlin Heidelberg: Springer.
- FAHRMEIR, L, KNEIB, T, & LANG, S. 2007. *Regression – Modelle, Methoden und Anwendungen*. Berlin Heidelberg: Springer.
- FANG, KT, & ZHANG, YT. 1990. *Generalized Multivariate Analysis*. Berlin Heidelberg: Springer.
- FRACKOWIAK, RSJ, FRISTON, KJ, FRITH, C, DOLAN, R, & MAZZIOTTA, JC. 1997. *Human Brain Function*. 1 edn. Academic Press.
- FRACKOWIAK, RSJ, FRISTON, KJ, FRITH, C, DOLAN, R, PRICE, CJ, ZEKI, S, ASHBURNER, JT, & PENNY, WD. 2003. *Human Brain Function*. 2 edn. Academic Press.
- FRISTON, KJ, PENNY, W, PHILLIPS, C, KIEBEL, S, HINTON, G, & ASHBURNER, J. 2002. Classical and Bayesian Inference in Neuroimaging: Theory. *NeuroImage*, **16**, 465–483.
- FRISTON, KJ, ASHBURNER, JT, KIEBEL, S, NICHOLS, TE, & PENNY, WD. 2007. *Statistical Parametric Mapping – The Analysis of Functional Brain Images*. Amsterdam: Academic Press.
- FUJIKOSHI, Y, ULYANOV, VV, & SHIMIZU, R. 2010. *Multivariate Statistics – High-Dimensional and Large-Sample Approximations*. Hoboken New Jersey: John Wiley & Sons.

- GENOVESE, CR, LAZAR, NA, & NICHOLS, T. 2001. Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage*, **15**, 870–878.
- GOOD, P. 2005. *Permutation, Parametric and Bootstrap Tests of Hypotheses*. 3 edn. New York: Springer.
- GRAY, RM. 2005. Toeplitz and Circulant Matrices: A review. *Foundations and Trends in Communications and Information Theory*, **2**(3), 155–329.
- HANDL, A. 2002. *Multivariate Analysemethoden: Theorie und Praxis multivariater Verfahren unter besonderer Berücksichtigung von S-PLUS*. Berlin Heidelberg: Springer.
- HARTUNG, J. 1986. *Statistik: Lehr- und Handbuch der angewandten Statistik*. 5 edn. München Wien: R. Oldenbourg Verlag.
- HASSABIS, D, CHU, C, REES, G, WEISKOPF, N, MOLYNEUX, PD, & MAGUIRE, EA. 2009. Decoding Neuronal Ensembles in the Human Hippocampus. *Current Biology*, **19**(7), 546–554.
- HOLLMANN, M. 2009. *Wissensrepräsentation zur integrierten Steuerung von Experimenten in der Echtzeit-fMRT: Neuroökonomische Untersuchung sozialer Interaktion*. Dissertation, Otto-von-Guericke-Universität Magdeburg.
- HOLLMANN, M, RIEGER, JM, BAECKE, S, LÜTZKENDORF, R, MÜLLER, C, ADOLF, D, & BERNARDING, J. 2011. Predicting Decisions in Human Social Interactions Using Real-Time fMRI and Pattern Classification. *Public Library of Science ONE*, **6**(10), e25304.
- HUIZENGA, HM, DEMUNCK, JC, WALDORP, LJ, & GRASMAN, RP. 2002. Spatio-temporal EEG/MEG Source Analysis Based on a Parametric Noise Covariance Model. *IEEE Transactions on Biomedical Engineering*, **49**(6), 533–539.
- HUMMEL, M, MEISTER, R, & MANSMANN, U. 2008. GlobalANCOVA: exploration and assessment of gene group effects. *Bioinformatics*, **24**(1), 78–85.
- JOLLIFFE, IT. 1972. Discarding Variables in a Principal Component Analysis II: Artificial data. *Applied Statistics*, **21**, 160–173.
- KAISER, HF. 1960. The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, **20**, 141–151.
- KHERAD-PAJOUH, S, & RENAUD, O. 2010. An exact permutation method for testing any effect in balanced and unbalanced fixed effect ANOVA. *Computational Statistics and Data Analysis*, **54**, 1881–1893.
- KIRCH, C. 2007. Block permutation principles for the change analysis of dependent data. *Journal of Statistical Planning and Inference*, **137**, 2453–2474.

- KLEIN, D, & ZEZULA, I. 2009. The maximum likelihood estimators in the growth curve model with serial covariance structure. *Journal of Statistical Planning and Inference*, **139**, 3270–3276.
- KOSE, D. 2007. *Verhalten multivariater Tests bei Abhängigkeiten zwischen den Stichprobenelementen sowie Maßnahmen zur Korrektur des Testniveaus mit Anwendung auf Daten aus der funktionellen Bildgebung*. Diplomarbeit, Hochschule Magdeburg-Stendal.
- KRIEGESKORTE, N, GOEBEL, R, & BANDETTINI, P. 2006. Information-based functional brain mapping. *PNAS*, **103**(10), 3863–3868.
- KROPF, S, & ADOLF, D. 2009. Rotation test with pairwise distance measures of sample vectors in a GLM. *Journal of Statistical Planning and Inference*, **139**, 3857–3864.
- KROPF, S, HEUER, H, GRÜNING, M, & SMALLA, K. 2004. Significance test for comparing complex microbial community fingerprints using pairwise similarity measures. *Journal of Microbiological Methods*, **57**(2), 187–195.
- KROPF, S, LUX, A, ESZLINGER, M, HEUER, H, & SMALLA, K. 2007. Comparison of Independent Samples of High-Dimensional Data by Pairwise Distance Measures. *Biometrical Journal*, **49**(2), 230–241.
- LANGSRUD, Ø. 2005. Rotation tests. *Statistics and Computing*, **15**, 53–60.
- LAUTERBUR, PC. 1973. Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature*, **242**(5394), 190–191.
- LAZAR, NA. 2008. *The Statistical Analysis of Functional MRI Data*. 2 edn. New York: Springer.
- LINDSEY, JK. 1994. *Models for Repeated Measurements*. Oxford: Clarendon Press.
- LUCHTMANN, M. 2010. *Einfluss von Ethanol auf das BOLD-Signal im motorischen und visuellen Kortex*. Dissertation, Otto-von-Guericke-Universität Magdeburg.
- LÄUTER, J. 2005. *Hochdimensionale Statistik – Anwendung in der Genexpressionsanalyse*. Interdisciplinary Centre for Bioinformatics. http://www.izbi.de/working_papers.html.
- LÄUTER, J, GLIMM, E, & KROPF, S. 1996. New Multivariate Tests for Data with an Inherent Structure. *Biometrical Journal*, **38**, 5–23.
- LÄUTER, J, GLIMM, E, & KROPF, S. 1998. Multivariate Tests Based on Left-Spherically Distributed Linear Scores. *Annals of Statistics*, **26**, 1972–1988.
- LÄUTER, J, GLIMM, E, & ESZLINGER, M. 2005. Search for relevant sets of variables in a high-dimensional setup keeping the familywise error rate. *Statistica Neerlandica*, **59**(3), 298–312.

- MANSFIELD, P, & GRANNELL, PK. 1973. NMR diffraction in solids. *Journal of Physics C: Solid State Physics*, **6**(22), L422–L426.
- MATLAB. 2011. *Matlab – The Language of Technical Computing*. <http://www.mathworks.de/products/matlab/>, zuletzt geprüft am 25.11.2011: The MathWorks, Inc.
- MEINSHAUSEN, N. 2008. Hierarchical testing of variable importance. *Biometrika*, **95**(2), 265–278.
- MIELKE, PW, & BERRY, KJ. 2007. *Permutation Methods – A Distance Function Approach*. 2 edn. New York: Springer.
- MIGNANI, S, & ROSA, R. 1995. The moving block bootstrap to assess the accuracy of statistical estimates in Ising model simulations. *Computer Physics Communications*, **92**, 203–213.
- NAUJOKS, S. 2008. *Verhalten von multivariaten Tests auf der Basis von paarweisen Ähnlichkeits- und Distanzmaßen zwischen den Stichprobenelementen bei Abhängigkeiten zwischen den Stichprobenelementen und Maßnahmen zur Korrektur des Testniveaus*. Diplomarbeit, Hochschule Magdeburg-Stendal.
- NICHOLS, T, & HAYASAKA, S. 2003. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, **12**, 419–446.
- OGAWA, S, LEE, TM, KAY, AR, & TANK, DW. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *PNAS*, **87**, 9868–9872.
- OHLSON, M, & KOSKI, T. 2009. *The Likelihood Ratio Statistic for Testing Spatial Independence using a Separable Covariance Matrix*. Technical Report LiTH-MAI-R-2009-06, Department of Mathematics, Linköping University.
- PENNY, WD, KILNER, J, & BLANKENBURG, F. 2007. Robust Bayesian general linear models. *NeuroImage*, **36**, 661–671.
- PESARIN, F, & SALMASO, L. 2010. *Permutation Tests for Complex Data: Theory, Applications and software*. Chichester: John Wiley & Sons.
- PHIPSON, B, & SMYTH, GK. 2010. Permutation P -values Should Never Be Zero: Calculating Exact P -values When Permutations Are Randomly Drawn. *Statistical Applications in Genetics and Molecular Biology*, **9**(1).
- PILLAI, KCS, & GUPTA, AK. 1969. On the exact distribution of Wilks' Criterion. *Biometrika*, **56**, 109–118.
- PIZIAK, R, & ODELL, PL. 2007. *Matrix Theory*. Boca Raton London New York: Chapman & Hall.

- POLITIS, DN, & ROMANO, JP. 1991 (February). *A circular block-resampling procedure for stationary data*. Technical report No. 91-07.
- PRUSCHA, H. 1996. *Angewandte Methoden der Mathematischen Statistik*. Stuttgart: B. G. Teubner.
- RASCH, D. 1978. *Einführung in die mathematische Statistik*. Berlin: Deutscher Verlag der Wissenschaften.
- SACHS, L. 2004. *Angewandte Statistik*. Berlin Heidelberg: Springer.
- SAS. 2008. *SAS 9.2 User's Guide, SAS Institute Inc*.
- SCHLITTGEN, R, & STREITBERG, HJ. 2001. *Zeitreihenanalyse*. 9 edn. München Wien: R. Oldenbourg Verlag.
- SCHNEIDER, F, & FINK, G. 2007. *Funktionelle MRT in Psychiatrie und Neurologie*. Heidelberg: Springer Medizin Verlag.
- SEBER, GAF. 1977. *Linear Regression Analysis*. New York Chichester Brisbane Toronto Singapore: John Wiley & Sons.
- SIEMENS. 2003. *Magnete, Spins und Resonanzen – Eine Einführung in die Grundlagen der Magnetresonanztomographie*. Erlangen: Siemens AG Medical Solutions.
- SOHR, M. 2007. *Analysis of Functional Magnetic Resonance Imaging Time Series by Independent Component Analysis*. Dissertation, Otto-von-Guericke-Universität Magdeburg.
- SPM. 2009. SPM[®] 8 Manual. <http://www.fil.ion.ucl.ac.uk/spm/>: Functional Imaging Laboratory, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, 12 Queen Square, London WC1N 3BG, UK.
- TALAIRACH, J, & TOURNOUX, P. 1988. *Co-Planar Stereotaxic Atlas of the Human Brain*. Thieme.
- WANG, H, & WEST, M. 2009. Bayesian analysis of matrix normal graphical models. *Biometrika*, **96**(4), 821–834.
- ZAINO, M. 2006. *Exakte parametrische multivariate Tests für hochdimensionale Beobachtungen mit Unterstützung der Auffindung faktorieller Strukturen*. Dissertation, Universität Leipzig.

Acknowledgements

Diese Arbeit wurde gefördert durch die Deutsche Forschungsgemeinschaft (DFG, Förderung: KR 2231/3-1/-2).

A. Totale Produktsummenmatrix bei Permutation

Grundlage der in dieser Arbeit verwendeten Permutationsstrategie ist das geteilte lineare Modell

$$\mathbf{Y} = (\mathbf{X}_1 \mathbf{X}_2) \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix} + \mathbf{E}, \quad \mathbf{E} \sim N(\mathbf{0}, \mathbf{I}_n \otimes \Sigma)$$

mit der korrespondierenden Nullhypothese $H_0 : \mathbf{B}_2 = \mathbf{0}_{s_2 \times p}$, das durch Transformation aus jedem allgemeinen linearen Modell gebildet werden kann (vgl. Abschnitt 2.2).

Als Teststatistik für die Permutation können beispielsweise stabilisierte multivariate Verfahren (vgl. Abschnitt 2.3.2) genutzt werden, bei denen die totale Produktsummenmatrix \mathbf{W} eine zentrale Rolle spielt. Sie ergibt sich aus

$$\mathbf{W} = \mathbf{H} + \mathbf{G} = \mathbf{Y}'(\mathbf{Q}_H + \mathbf{Q}_G)\mathbf{Y} = \mathbf{Y}'\mathbf{Q}_W\mathbf{Y}$$

mit

$$\mathbf{Q}_H = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C} \left(\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C} \right)^{-1} \mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

und

$$\mathbf{Q}_G = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

(vgl. Satz 2.7). Im Falle des geteilten Modells sind hierbei $\mathbf{C} = \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{s_2} \end{pmatrix}$ und $\mathbf{X} = (\mathbf{X}_1 \mathbf{X}_2)$, wobei ohne Permutation einer Designteilmatrix $\mathbf{X}'_1\mathbf{X}_2 = \mathbf{X}'_2\mathbf{X}_1 = \mathbf{0}$ gilt.

Weiterhin ist die für \mathbf{Q}_W benötigte Inverse

$$\begin{aligned} (\mathbf{X}'\mathbf{X})^{-1} &= \left(\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} (\mathbf{X}_1 \mathbf{X}_2) \right)^{-1} \\ &= \begin{pmatrix} \mathbf{A} & -(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{X}_2\mathbf{B} \\ -(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{X}_1\mathbf{A} & \mathbf{B} \end{pmatrix} \end{aligned} \quad (\text{A.1})$$

mit

$$\mathbf{A} = \left(\mathbf{X}'_1\mathbf{X}_1 - \mathbf{X}'_1\mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{X}_1 \right)^{-1} \quad (\text{A.2})$$

und

$$\mathbf{B} = \left(\mathbf{X}'_2\mathbf{X}_2 - \mathbf{X}'_2\mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{X}_2 \right)^{-1}$$

(vgl. Piziak & Odell, 2007).

Im orthogonalen Fall vereinfacht sich diese Inverse zu

$$\begin{pmatrix} (\mathbf{X}'_1 \mathbf{X}_1)^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \end{pmatrix}.$$

Generell ergibt sich mit (A.1) \mathbf{Q}_W aus

$$\begin{aligned} \mathbf{Q}_W &= (\mathbf{X}_1 \mathbf{X}_2) \begin{pmatrix} \mathbf{A} & -(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 \mathbf{B} \\ -(\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{X}_1 \mathbf{A} & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix} \\ &\quad \left[(\mathbf{0} \ \mathbf{I}) \begin{pmatrix} \mathbf{A} & -(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 \mathbf{B} \\ -(\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{X}_1 \mathbf{A} & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix} \right]^{-1} \\ &\quad (\mathbf{0} \ \mathbf{I}) \begin{pmatrix} \mathbf{A} & -(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 \mathbf{B} \\ -(\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{X}_1 \mathbf{A} & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{X}'_1 \\ \mathbf{X}'_2 \end{pmatrix} \\ &+ \mathbf{I}_n - \left[(\mathbf{X}_1 \mathbf{X}_2) \begin{pmatrix} \mathbf{A} & -(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 \mathbf{B} \\ -(\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{X}_1 \mathbf{A} & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{X}'_1 \\ \mathbf{X}'_2 \end{pmatrix} \right]. \quad (\text{A.3}) \end{aligned}$$

Dies kann ausmultipliziert werden zu

$$\begin{aligned} \mathbf{Q}_W &= \left(-\mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 \mathbf{B} + \mathbf{X}_2 \mathbf{B} \right) \mathbf{B}^{-1} \left(-(\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{X}_1 \mathbf{A} \mathbf{X}'_1 + \mathbf{B} \mathbf{X}'_2 \right) \\ &\quad + \mathbf{I}_n - \left(\mathbf{I}_n - \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 \right) \mathbf{X}_1 \mathbf{A} \mathbf{X}'_1 - \left(\mathbf{I}_n - \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \right) \mathbf{X}_2 \mathbf{B} \mathbf{X}'_2. \end{aligned}$$

Weiterhin kann dieser Ausdruck zusammengefasst werden zu

$$\mathbf{Q}_W = \mathbf{I}_n + \left(\mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 - \mathbf{I}_n \right) \mathbf{X}_1 \mathbf{A} \mathbf{X}'_1.$$

Ersetzt man \mathbf{A} wieder durch (A.2), ergibt sich

$$\begin{aligned} \mathbf{Q}_W &= \mathbf{I}_n + \left(\mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 - \mathbf{I}_n \right) \\ &\quad \mathbf{X}_1 \left(\mathbf{X}'_1 \mathbf{X}_1 - \mathbf{X}'_1 \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{X}_1 \right)^{-1} \mathbf{X}'_1. \end{aligned}$$

Dies lässt sich wiederum umformulieren zu

$$\begin{aligned} \mathbf{Q}_W &= \mathbf{I}_n - \left(-\mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 + \mathbf{I}_n \right) \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-\frac{1}{2}} \\ &\quad \left(\mathbf{I}_n - (\mathbf{X}'_1 \mathbf{X}_1)^{-\frac{1}{2}} \mathbf{X}'_1 \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-\frac{1}{2}} \right)^{-1} (\mathbf{X}'_1 \mathbf{X}_1)^{-\frac{1}{2}} \mathbf{X}'_1, \quad (\text{A.4}) \end{aligned}$$

so dass auch der erste Summand innerhalb der Inversen invertierbar ist. Ersetzt man $\mathbf{D}_1 = \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-\frac{1}{2}}$ und $\mathbf{D}_2 = \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{X}_2)^{-\frac{1}{2}}$ wird (A.4) zu

$$\mathbf{Q}_W = \mathbf{I}_n - \left(-\mathbf{D}_1 \mathbf{D}'_1 \mathbf{D}_2 \mathbf{D}'_2 + \mathbf{I}_n \right) \mathbf{D}_1 \left(\mathbf{I}_n - \mathbf{D}'_1 \mathbf{D}_2 \mathbf{D}'_2 \mathbf{D}_1 \right)^{-1} \mathbf{D}'_1.$$

Die Inverse kann nach Ahrens & Läuter (1981) umgeformt werden, so dass man

$$\mathbf{Q}_W = \mathbf{I}_n - \left(-\mathbf{D}_1 \mathbf{D}'_1 \mathbf{D}_2 \mathbf{D}'_2 + \mathbf{I}_n \right) \mathbf{D}_1 \left(\mathbf{I}_n + \mathbf{D}'_1 \mathbf{D}_2 (\mathbf{I}_n - \mathbf{D}'_2 \mathbf{D}_1 \mathbf{D}'_1 \mathbf{D}_2)^{-1} \mathbf{D}'_2 \mathbf{D}_1 \right) \mathbf{D}'_1$$

erhält. Durch Ausmultiplizieren und Umstellen entsteht daraus

$$\begin{aligned} \mathbf{Q}_W &= \mathbf{I}_n - \mathbf{D}_1 \mathbf{D}'_1 \\ &\quad + \mathbf{D}_1 \mathbf{D}'_1 \mathbf{D}_2 \left[-(\mathbf{I}_n - \mathbf{D}'_2 \mathbf{D}_1 \mathbf{D}'_1 \mathbf{D}_2)^{-1} + \mathbf{I}_n \right. \\ &\quad \left. + \mathbf{D}'_2 \mathbf{D}_1 \mathbf{D}'_1 \mathbf{D}_2 (\mathbf{I}_n - \mathbf{D}'_2 \mathbf{D}_1 \mathbf{D}'_1 \mathbf{D}_2)^{-1} \right] \mathbf{D}'_2 \mathbf{D}_1 \mathbf{D}'_1. \end{aligned}$$

Dabei entsprechen die Summanden in den letzten beiden Zeilen einer quadratischen Form, deren in eckige Klammern gefasster Teil eine Nullmatrix ergibt, so dass letztendlich für (A.3) $\mathbf{Q}_W = \mathbf{I}_n - \mathbf{D}_1 \mathbf{D}'_1$, also

$$\mathbf{Q}_W = \mathbf{I}_n - \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1$$

übrig bleibt.

Daher ist die totale Produktsummenmatrix \mathbf{W} vollständig von der Designteilmatrix \mathbf{X}_2 unabhängig.

B. Erwartungswert des AR(1)-Korrelationskoeffizientenschätzers

Der Erwartungswert des geschätzten Korrelationskoeffizienten eines autoregressiven Prozesses erster Ordnung in [Schlittgen & Streitberg \(2001\)](#) ist

$$E(\hat{\rho}) = \rho - \frac{1}{n} \left[\rho + \frac{f(0)}{\gamma_0} (1 - \rho) + 2\zeta_1 - 2\rho\zeta_0 \right] + O(n^{-2}) \quad (\text{B.1})$$

mit

$$\zeta_c = \frac{\rho^{|c|}}{1 - \rho^2} (1 + \rho^2 + |c|(1 - \rho^2)),$$

so dass

$$\zeta_0 = \frac{1 + \rho^2}{1 - \rho^2} \quad \text{and} \quad \zeta_1 = \frac{2\rho}{1 - \rho^2}.$$

Die Kovarianzfunktion eines AR(1) Prozesses ist

$$\gamma_\tau = \rho^{|\tau|} \frac{\sigma_e^2}{1 - \rho^2},$$

wobei der sogenannte Lag $\tau = s - t$ für den Abstand der Zeitpunkte steht. Für $\tau = 0$ ergibt sich

$$\gamma_0 = \frac{\sigma_e^2}{1 - \rho^2}.$$

Das Spektrum $f(0)$, die Fourier-Transformation der Kovarianzfunktion, ist

$$\begin{aligned} f(0) &= \frac{\sigma_e^2}{1 + \rho^2 - 2\rho \cos 2\pi 0} \\ &= \frac{\sigma_e^2}{1 + \rho^2 - 2\rho}. \end{aligned}$$

Demzufolge kann der Erwartungswert für die Schätzung des AR(1) Korrelationskoeffizienten (B.1) vereinfacht werden zu

$$\begin{aligned} E(\hat{\rho}) &= \rho - \frac{1}{n} \left[\rho + \frac{1 - \rho^2}{1 + \rho^2 - 2\rho} (1 - \rho) + 2 \frac{2\rho}{1 - \rho^2} - 2\rho \frac{1 + \rho^2}{1 - \rho^2} \right] + O(n^{-2}) \\ &= \rho - \frac{1}{n} [1 + 4\rho] + O(n^{-2}). \end{aligned}$$

C. Weitere Simulationsergebnisse für den Fehler 1. Art

Design	Blocklänge	ρ				
		0,0	0,2	0,4	0,6	0,8
Block	1	0,050	0,134	0,280	0,430	0,580
	5	0,048	0,066	0,099	0,146	0,236
	10	0,047	0,059	0,068	0,081	0,132
	15	0,044	0,052	0,063	0,061	0,092
	20	0,045	0,050	0,064	0,057	0,082
	40	0,044	0,049	0,055	0,049	0,065
	60	0,046	0,047	0,050	0,045	0,053
Event	1	0,050	0,113	0,169	0,219	0,268
	5	0,046	0,062	0,056	0,071	0,075
	10	0,047	0,057	0,051	0,062	0,066
	15	0,047	0,052	0,050	0,056	0,062
	20	0,045	0,056	0,046	0,057	0,058
	40	0,052	0,051	0,043	0,053	0,054
	60	0,049	0,049	0,040	0,057	0,050

Tabelle C.1.: Empirischer Fehler 1. Art der Blockpermutation mit Random Shift auf Basis der eindimensionalen skaleninvarianten Principal Component Teststatistik bei simulierten fMRT-Daten für unterschiedliche fMRT-Designs ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

Design	Blocklänge	ρ				
		0,0	0,2	0,4	0,6	0,8
Block	1	0,045	0,379	0,985	1,000	1,000
	5	0,046	0,089	0,426	1,000	1,000
	10	0,049	0,071	0,192	0,826	1,000
	15	0,047	0,062	0,119	0,499	0,897
	20	0,048	0,057	0,099	0,362	0,800
	40	0,048	0,057	0,067	0,150	0,284
	60	0,047	0,052	0,056	0,091	0,100
Event	1	0,050	0,266	0,792	0,999	1,000
	5	0,043	0,080	0,114	0,218	0,155
	10	0,046	0,062	0,066	0,117	0,084
	15	0,045	0,057	0,061	0,090	0,068
	20	0,046	0,054	0,054	0,073	0,055
	40	0,046	0,054	0,048	0,068	0,057
	60	0,046	0,049	0,046	0,067	0,062

Tabelle C.2.: Empirischer Fehler 1. Art der Blockpermutation mit Random Shift auf Basis der q -dimensionalen skaleninvarianten Principal Component Teststatistik bei simulierten fMRT-Daten für unterschiedliche fMRT-Designs ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

		ρ				
Design	Blocklänge	0,0	0,2	0,4	0,6	0,8
Block	1	0,047	0,372	0,827	0,994	1,000
	5	0,046	0,090	0,169	0,371	0,765
	10	0,046	0,066	0,106	0,160	0,327
	15	0,046	0,059	0,085	0,107	0,171
	20	0,042	0,057	0,072	0,092	0,124
	40	0,048	0,053	0,060	0,065	0,073
	60	0,046	0,049	0,057	0,058	0,064
Event	1	0,047	0,250	0,561	0,773	0,866
	5	0,046	0,065	0,102	0,112	0,086
	10	0,047	0,054	0,077	0,073	0,069
	15	0,046	0,050	0,068	0,064	0,067
	20	0,051	0,049	0,063	0,064	0,062
	40	0,051	0,042	0,057	0,054	0,056
	60	0,046	0,044	0,056	0,052	0,056

Tabelle C.3.: Empirischer Fehler 1. Art der Blockpermutation mit Random Shift auf Basis der modifizierten q -dimensionalen skaleninvarianten Principal Component Teststatistik bei simulierten fMRT-Daten für unterschiedliche fMRT-Designs ($n = 250$, $p = 1024$, 299 Permutationen, 2500 Wiederholungen)

		ρ				
Design	Adjustierung	0,0	0,2	0,4	0,6	0,8
Block	Original	0,054	0,140	0,288	0,441	0,579
	Satterthwaite	0,056	0,044	0,017	0,001	0,000
	Pre-Whitening	0,056	0,053	0,056	0,062	0,061
	Blockpermutation	0,046	0,047	0,050	0,045	0,053
Event	Original	0,054	0,116	0,177	0,231	0,279
	Satterthwaite	0,056	0,043	0,012	0,003	0,000
	Pre-Whitening	0,055	0,059	0,048	0,060	0,049
	Blockpermutation	0,049	0,049	0,040	0,057	0,050

Tabelle C.4.: Empirischer Fehler 1. Art des eindimensionalen skaleninvarianten Principal Component Tests bei simulierten fMRT-Daten für verschiedene Adjustierungsverfahren und fMRT-Designs ($n = 250$, $p = 1024$, 299 Permutationen, $l = 60$, 2500 Wiederholungen)

		ρ				
Design	Adjustierung	0,0	0,2	0,4	0,6	0,8
Block	Original	0,050	0,393	0,986	1,000	1,000
	Satterthwaite	0,055	0,035	0,000	0,000	0,000
	Pre-Whitening	0,055	0,062	0,068	0,069	0,073
	Blockpermutation	0,047	0,052	0,056	0,091	0,100
Event	Original	0,048	0,275	0,804	0,999	1,000
	Satterthwaite	0,052	0,038	0,000	0,000	0,000
	Pre-Whitening	0,052	0,056	0,050	0,052	0,044
	Blockpermutation	0,046	0,049	0,046	0,067	0,062

Tabelle C.5.: Empirischer Fehler 1. Art des q -dimensionalen skaleninvarianten Principal Component Tests bei simulierten fMRT-Daten für verschiedene Adjustierungsverfahren und fMRT-Designs ($n = 250$, $p = 1024$, 299 Permutationen, $l = 60$, 2500 Wiederholungen)

		ρ				
Design	Adjustierung	0,0	0,2	0,4	0,6	0,8
Block	Original	0,049	0,364	0,814	0,996	1,000
	Pre-Whitening	0,064	0,068	0,076	0,081	0,082
	Blockpermutation	0,046	0,049	0,057	0,058	0,064
Event	Original	0,054	0,244	0,531	0,776	0,866
	Pre-Whitening	0,060	0,052	0,063	0,060	0,050
	Blockpermutation	0,046	0,044	0,056	0,052	0,056

Tabelle C.6.: Empirischer Fehler 1. Art des modifizierten q -dimensionalen skaleninvarianten Principal Component Tests bei simulierten fMRT-Daten für verschiedene Adjustierungsverfahren und fMRT-Designs ($n = 250$, $p = 1024$, 299 Rotationen bzw. Permutationen, $l = 60$, 2500 Wiederholungen)

Design	Adjustierung	PC(1)	PC(q)	mod. PC(q)
Block	Original	0,446	1,000	1,000
	Pre-Whitening	0,063	0,122	0,126
	Blockpermutation	0,052	0,081	0,072
Event	Original	0,230	1,000	1,000
	Pre-Whitening	0,060	0,069	0,070
	Blockpermutation	0,045	0,068	0,071

Tabelle C.7.: Empirischer Fehler 1. Art der skaleninvarianten Principal Component Teststatistiken bei simulierten fMRT-Daten mit zufälliger räumlicher Korrelation für verschiedene Adjustierungsverfahren und fMRT-Designs ($n = 250$, $p = 1024$, $\rho = 0, 4$, 299 Rotationen bzw. Permutationen, $l = 60$, 2500 Wiederholungen)

Design	Adjustierung	PC(1)	PC(q)	mod. PC(q)
Block	Original	0,470	0,990	1,000
	Pre-Whitening	0,125	0,226	0,210
	Blockpermutation	0,092	0,185	0,133
Event	Original	0,252	0,790	0,638
	Pre-Whitening	0,075	0,101	0,102
	Blockpermutation	0,064	0,077	0,063

Tabelle C.8.: Empirischer Fehler 1. Art der skaleninvarianten Principal Component Teststatistiken bei simulierten fMRT-Daten für verschiedene Adjustierungsverfahren und fMRT-Designs bei verkleinertem Stichprobenumfang ($n = 50$, $p = 1024$, $\rho = 0, 4$, 299 Rotationen bzw. Permutationen, $l = 10$, 2500 Wiederholungen)

D. Weitere Simulationsergebnisse für die Güte

		ρ				
Design	Adjustierung	0,0	0,2	0,4	0,6	0,8
Block	Satterthwaite	0,606	0,410	0,184	0,018	0,000
	Pre-Whitening	0,607	0,463	0,334	0,246	0,180
	Blockpermutation	0,570	0,513	0,430	0,314	0,184
Event	Satterthwaite	0,489	0,351	0,164	0,032	0,000
	Pre-Whitening	0,493	0,386	0,330	0,336	0,506
	Blockpermutation	0,465	0,424	0,331	0,244	0,152

Tabelle D.1.: Empirische Güte der verschiedenen Adjustierungsverfahren des eindimensionalen skaleninvarianten Principal Component Tests bei simulierten fMRT-Daten für beide fMRT-Designs ($n = 250$, $p = 1024$, 128 Variablen unter der Alternative, 299 Permutationen, $l = 60$, 2500 Wiederholungen)

		ρ				
Design	Adjustierung	0,0	0,2	0,4	0,6	0,8
Block	Satterthwaite	0,999	0,939	0,241	0,000	0,000
	Pre-Whitening	0,999	0,964	0,823	0,608	0,478
	Blockpermutation	0,991	0,938	0,322	0,043	0,107
Event	Satterthwaite	0,971	0,844	0,215	0,000	0,000
	Pre-Whitening	0,972	0,899	0,820	0,835	0,982
	Blockpermutation	0,960	0,878	0,614	0,479	0,704

Tabelle D.2.: Empirische Güte der verschiedenen Adjustierungsverfahren des q -dimensionalen skaleninvarianten Principal Component Tests bei simulierten fMRT-Daten für beide fMRT-Designs ($n = 250$, $p = 1024$, 128 Variablen unter der Alternative, 299 Permutationen, $l = 60$, 2500 Wiederholungen)

Design	Adjustierung	ρ				
		0,0	0,2	0,4	0,6	0,8
Block	Pre-Whitening	0,999	0,963	0,828	0,610	0,452
	Blockpermutation	0,995	0,949	0,773	0,510	0,294
Event	Pre-Whitening	0,976	0,906	0,830	0,850	0,981
	Blockpermutation	0,965	0,891	0,770	0,606	0,429

Tabelle D.3.: Empirische Güte der verschiedenen Adjustierungsverfahren des modifizierten q -dimensionalen skaleninvarianten Principal Component Tests bei simulierten fMRT-Daten für beide fMRT-Designs ($n = 250$, $p = 1024$, 128 Variablen unter der Alternative, 299 Rotationen bzw. Permutationen, $l = 60$, 2500 Wiederholungen)

Design	Adjustierung	PC(1)	PC(q)	mod. PC(q)
Block	Pre-Whitening	0,175	0,460	0,459
	Blockpermutation	0,193	0,199	0,421
Event	Pre-Whitening	0,168	0,457	0,465
	Blockpermutation	0,162	0,285	0,393

Tabelle D.4.: Empirische Güte der skaleninvarianten Principal Component Teststatistiken bei simulierten fMRT-Daten für verschiedene Adjustierungsverfahren und fMRT-Designs mit 64 Variablen unter der Alternative ($n = 250$, $p = 1024$, $\rho = 0,4$, 64 Variablen unter der Alternative, 299 Rotationen bzw. Permutationen, $l = 60$, 2500 Wiederholungen)

Design	Adjustierung	PC(1)	PC(q)	mod. PC(q)
Block	Pre-Whitening	0,208	0,993	0,990
	Blockpermutation	0,578	0,989	0,996
Event	Pre-Whitening	0,212	0,990	0,983
	Blockpermutation	0,250	0,994	0,991

Tabelle D.5.: Empirische Güte der skaleninvarianten Principal Component Teststatistiken bei simulierten fMRT-Daten mit zufälliger Korrelation für verschiedene Adjustierungsverfahren und fMRT-Designs ($n = 250$, $p = 1024$, $\rho = 0,4$, 64 Variablen unter der Alternative, 299 Rotationen bzw. Permutationen, $l = 60$, 2500 Wiederholungen)

Danksagung

*Was wir wissen, ist ein Tropfen;
was wir nicht wissen, ist ein Ozean.*

Isaac Newton

In erster Linie möchte ich Prof. Dr. Siegfried Kropf für seine umfassende Unterstützung bei der Erstellung dieser Dissertation danken. Seine wegweisenden Ratschläge und Anregungen haben entscheidend zum Gelingen dieser Arbeit beigetragen. Zudem weckte er meine Begeisterung für das Gebiet der medizinischen Statistik und wies mir den Weg in die wissenschaftliche Gemeinschaft.

Mein herzlicher Dank gilt im Weiteren Prof. Dr. Waltraud Kahle für die Betreuung der Promotion, ihre fortwährende Ermutigung und zahlreiche Hinweise bei der Bearbeitung.

Diese Dissertation entstand innerhalb des DFG-Forschungsprojektes “Multivariate und multiple Testverfahren für hochdimensionale Daten bei zeitlich abhängigen Beobachtungen mit Anwendungen auf fMRI-Daten” am Institut für Biometrie und Medizinische Informatik der Otto-von-Guericke-Universität Magdeburg unter Leitung von Prof. Dr. Dr. Johannes Bernarding. Er und alle Kolleginnen und Kollegen boten mir während dieser Zeit ein angenehmes Arbeitsumfeld, wofür ich mich herzlich bedanken möchte. Ein besonderer Dank gilt dabei Dr. Friedrich-Wilhelm Röhl für seine beharrliche Motivation, sowie Snezhana Weston, die mir bei wesentlichen Details oft wertvolle Hilfestellungen leistete. Weiterhin möchte ich mich bei den Mitarbeitern der Medizinischen Informatik, insbesondere Sebastian Baecke und Charles Müller, sowie Dr. Maurice Hollmann bedanken, die mir stets in Bezug auf die funktionelle Magnetresonanztomographie mit Rat, Tat und Daten zur Seite standen.

Schließlich danke ich von ganzem Herzen meiner Familie und meinen Freunden für ihre andauernde Unterstützung vor und während der Promotion sowie meinem Mann für seine immerwährende Motivation und seinen liebevollen Beistand.